OBITUARY NOTICE

GEORGE UDNY YULE, C.B.E., F.R.S.

With the passing of George Udny Yule, psychological investigators lost one who had been a sympathetic statistical adviser during four decades. As early as 1911 he collaborated with the editors of the *British Journal of Psychology* in the selection of papers for publication, and he served on the Statistical Committee of the Industrial Fatigue Research Board (later the Industrial Health Research Board) almost from its inception, himself contributing a report on the function of statistical method in scientific investigation. Until 1930–1 he gave courses of lectures in statistical methods to students in the Psychological Laboratory at Cambridge and guided the steps of many future research workers in the right direction. Further, his work with the late Major Greenwood on the incidence and distribution of accidents (published in 1920) led directly to the vast amount of research into the causation of accidents which has since been undertaken.

Born in 1871, Yule was educated at Winchester and University College, London. At the age of 22 he was appointed a lecturer in the new school of mathematical statistics founded by Karl Pearson at University College. In the following years he published a number of papers on the theory of correlation and association—papers which are still quoted and referred to by serious students of statistics. His famous text-book, *An Introduction to the Theory of Statistics*, now in its fourteenth edition, was first published in 1911. In 1912 Yule went to Cambridge, where he remained until his death in 1951. For the major portion of this period he held the position of University Reader in Statistics, living in St John's College as a Fellow. His considerable original research work led to his election as a Fellow of the Royal Society in 1921, and in 1925 the Royal Statistical Society elected him President.

On his retirement from the Readership at the age of 60, Yule sought fresh outlets for his abundant energies. He learned to fly and obtained a pilot's licence and he embarked upon statistical researches involving a tremendous amount of detailed work—his capacity for taking pains tended towards infinity. These later studies were in the literary field and were described in *The Statistical Study of Literary Vocabulary*, published in 1944. As had so often been the case in his earlier work, Yule was a pioneer in this field and his remarkable book contained *inter alia* an interesting application of the theory of accident incidence to the use of words by different authors.

Yule was essentially a modest man, with a delightful sense of humour and great friendliness in his dealings with others. So many statistical advisers tend to lose sight of the nature of the material presented for analysis and are apt to dictate to the unfortunate experimentalist who consults them. Not so Yule, however. He insisted that the better a biological experiment, the less the need for statistics. Whilst laying no claim to expert knowledge in psychological and kindred fields—indeed, typically, he even denied being a mathematician—yet he was always available to give clear and sound advice in his own field of studies, a field in which his name will for ever live as one of the great ones.

E. G. C.

THE 'PSYCHOLOGICAL REFRACTORY PERIOD' AND THE TIMING OF HIGH-SPEED PERFORMANCE— A REVIEW AND A THEORY

By A. T. WELFORD

Nuffield Research Unit into Problems of Ageing, The Psychological Laboratory, University of Cambridge

I. *Introduction and statement of a theory* (pp. 1–5). II. *Survey and discussion of the evidence:* (i) *Two stimuli presented close together* (pp. 5–7); (ii) *Grouping of stimuli* (pp. 8–9); (iii) *Stopping a response* (pp. 9–12); (iv) *Amending a response* (pp. 12–14); (v) *Continuous performance at tracking tasks, etc* (pp. 14–18). III. *Summary* (pp. 18–19). *References* (p. 19).

I. Introduction and statement of a theory

The concept of the so-called psychological refractory period has arisen from studies o hand reaction times to serial stimuli and from observations that corrections in tracking tasks are apparently not continuous but intermittent. When discrete stimuli are presented in a series at irregular intervals the reaction time to the second of two stimuli separated by an interval of about 0·5 sec. or less tends to be longer than to the first. The explanations given of this phenomenon may be divided into three classes:

(1) The first stimulus is held to render the responding mechanism insensitive in the sense that it *acts more slowly* in dealing with a second stimulus arriving within a short interval.

(2) The delay of the second response is regarded as being due to the second stimulus being *unexpected.*

(3) The delay is held to be due to *some phase of the two reactions not being able to overlap.* Some phase of the first reaction is not finished by the time the second stimulus arrives, so that there is a delay before the second reaction can begin.

What is the nature of the insensitivity postulated by (1) is not entirely clear, but it would appear to be conceived as being great during the period immediately following the passing of the data from a stimulus to the central mechanisms and as gradually dying away so that when about 0·5 sec. has elapsed it has become very small. An analogy between it and the refractory period in nerves—with, of course, a difference of time scale—has been drawn, and it is due to this that the term 'psychological refractory period' has come into use. The analogy is a loose one because refractoriness of nerves means the requirement of a stronger signal, whereas 'psychological refractoriness' means increased latency. Such a concept would seem to fit well with much—perhaps all—of the published data, but has the disadvantage of postulating an *ad hoc* functional 'entity' and thus of being descriptive rather than explanatory.

With regard to (2), it is well known that reaction times to unexpected stimuli tend to be longer than to expected, and the experiments of Telford (12), Hick (4) and Vince (13) have been compared with that of Mowrer (8). Mowrer found that subjects trained to react to stimuli at 12 sec. intervals showed reaction times which were slightly longer to occasional stimuli coming at longer intervals, and very much longer to occasional stimuli at shorter intervals. This he explained as being due to 'low expectancy' of intervals other than 12 sec.

The similarity to Mowrer's results would, however, seem to be superficial, because in his experiment the short intervals were very rare, whereas in the other experiments they were relatively common. Even if the explanation in terms of expectancy is partly correct, the problem is merely shifted to the question: 'In what does expectancy consist?' In Mowrer's and the classical reaction time experiments, it seems to have meant that the subject was able to predict the arrival of the stimulus and was thus able either to prepare for it in the sense of doing some of the 'work' of reacting before it appeared, or at least to ensure that when it arrived he was not doing something else which would prevent him from giving it immediate attention.

Now, if the analogy with Mowrer's experiment were correct, and if preparation were something which could go on simultaneously with other activities such as reacting to a previous stimulus, there would seem to be no reason why refractoriness could not be eliminated with suitable training. Hick (4) points out, however, that refractoriness persists in the face of considerable practice in which attempts are made to reduce it, and it would therefore appear that the observed delays are compulsive, and presumably due to interference by some phase of the previous reaction. If this is so, it means that when intervals between stimuli are very short, lack of preparation would appear not to be capable of explaining refractoriness; rather, refractoriness of some kind is required to explain lack of preparation.

If we attempt an explanation of the delay by postulating that some phase of the first reaction is not finished, our task is essentially to answer the question: 'What information do we need and what postulates must we make about the first reaction in order to predict the reaction time to the second?'

The delay is clearly due sometimes to the time taken to execute the first responding action, but this is not always so. Not only does delay appear when the observed movement time is too short to account for it, but if this were the limiting factor we should not expect delay to occur when successive responses are made with different muscles. Some evidence has been provided by Hick that delay is still present when successive responses are made with different fingers (6).

It is also clear that the delay is not due to the central mechanisms being unable to prepare a fresh action while a former action is being carried out. In certain circumstances, at least, the central mechanisms can prepare a new response while a previous response is in progress (see (16)).

The evidence would seem clearly to indicate that the refractoriness is in the central mechanisms themselves, and the theory has been put forward that it is due to the central processes concerned with two separate stimuli not being able to co-exist, so that the data from a stimulus which arrives while the central mechanisms are dealing with data from a previous stimulus have to be 'held in store' until the mechanisms have been cleared.

According to this hypothesis

$$T_{R_2} = T_{R_1} + T_{O_2} - I, \tag{1}$$

when T_R = reaction time, i.e. the time between the onset of a stimulus and the beginning of the response movement;

T_O = time required by the central mechanisms to deal with the data provided by a stimulus and to initiate a response. We may call this the *central organizing time*. Normally it will be the same as T_R for a single response.

I = time interval between two stimuli S_1 and S_2.

If we assume that T_O is the same for all stimuli whether or not they follow closely after others, i.e. approximately $= T_{R_1}$, this formula accounts well, within statistical limits, for the delays observed when $I < T_{R_1}$. It does not, however, adequately account for the delays observed when $I > T_{R_1}$ (e.g. Vince (13, 16)).

Attempts appear sometimes to have been made to get round this difficulty by the assumption that the central mechanisms remain 'busy' for an appreciable period after the response has begun. This, however, is unsatisfactory because the delay it postulates is too long to match the observed reaction times when $I < T_{R_1}$.

What appears to be needed is the assumption of absolute refractoriness during the first reaction time, and conditional refractoriness thereafter. Hick ((4), p. 43) has suggested a source of such conditional refractoriness in attention by the subject to the response movement. He says: 'the attention may be switched to that sensory field from which confirmation of the occurrence of the response is expected. Or alternatively, to avoid the teleological concept of "expecting confirmation", we may suppose that the attention is reflexly deflected by the inevitable stimulation of kinaesthetic or other receptors by the response. Now it seems very likely that a second stimulus, closely following the first, demands, as its initial response, the switching of the attention back to it, and it may be this that occupies the extra time.' This important suggestion seems to need supplementing in one respect. If the stimuli fed back from a response are dealt with by the subject's central mechanisms in a similar way to stimuli given by the experimenter, we should expect that those fed back at the *beginning* of the response would 'capture' the central mechanisms for a brief period, and that only when this period was over could any further stimuli from the response, or any new stimulus given by the experimenter, be attended to. In short, perception of the stimuli resulting from the response would be intermittent and would give rise to one or more periods of refractoriness during the response movement. One such period would seem especially liable to occur immediately after the beginning of the response movement, but others might occur, especially where 'high points' of kinaesthetic and other stimulation arise. Where these 'high points' occur will vary according to the nature of the movement. Thus, when a rapid graded movement is made from one position to another the initial acceleration and final deceleration would seem likely to provide more intense sensory impulses than the middle portion of the movement. When a morse key is pressed, sensory impulses likely to 'capture' the central mechanisms might be expected at the beginning of the movement, when the bottom stop is struck and again when the hand is brought to rest after the key has been released. The possible occurrence of a 'high point' of stimulation at the end of a movement has the important implication that a short refractory period may *follow* the making of a response and cause some delay in the response to a second stimulus given by the experimenter even though this occurs after the previous movement has been completed.

With M = response movement and F = sensory data or stimulus 'fed back' from a response, let us write

T_{Fb} and T_{Fe} = times for organizing sensory data fed back from the beginning and the end respectively of a response movement.

T_M = time occupied in making a response movement.

Then on the view stated we should expect:

When $T_{R_1} < I < (T_{R_1} + T_{Fb_1})$

$$T_{R_2} = T_{R_1} + T_{Fb_1} + T_{O_2} - I, \qquad (3$$

and when $(T_{R_1}+T_{M_1})<I<(T_{R_1}+T_{M_1}+T_{Fe_1})$

$$T_{R_2}=T_{R_1}+T_{M_1}+T_{Fe_1}+T_{O_2}-I. \quad (3)$$

The durations of T_{Fb} and T_{Fe} are likely to vary with the amount of information *Fb* and *Fe* convey, but it appears from the experimental evidence that if T_{Fb} and T_{Fe} are taken as 0·15 sec. each, these two formulae can account reasonably well, within statistical limits, for the observed T_{R_2} when $I>T_{R_1}$.

II. Survey and discussion of the evidence

(i) *Two stimuli presented close together*

a) Craik (2) *and Vince* (13)

The subject's task in this experiment was to keep a pencil on a line drawn on a horizontally moving band of paper exposed to view behind a narrow vertical slit. From time to time the line moved suddenly upwards or downwards a distance of 2·5 cm. In the course of the task a number of paired stimuli, consisting of a step in one direction closely followed by a step back to the original level, was presented with $I=0{\cdot}05$, 0·1, 0·2, 0·3, 0·4, 0·5, 0·6, 0·7, 1·2 and 1·6 sec. It was clear from the results that at least in the cases where I was less than 0·4 sec., T_{R_2} tended to be lengthened. The results according to the theory proposed are set out for single stimuli in Fig. 1, and for pairs of stimuli in Fig. 2, using the following quantities:

T_{R_1} = the observed mean first reaction times;
$T_{O_2}=0{\cdot}29$ sec. = the observed mean T_R for first and single stimuli;
$T_M=0{\cdot}22$ sec. = the observed mean T_M for single stimuli;
$T_{Fb}=T_{Fe}=0{\cdot}15$ sec.

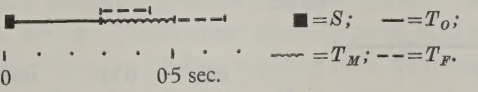


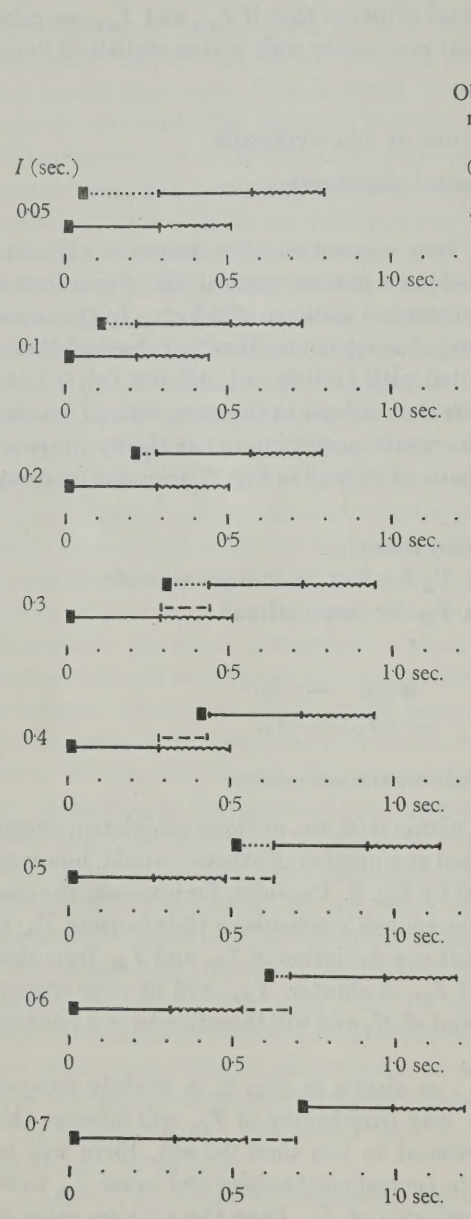
Fig. 1. Time components of single stimulus and response.

It will be seen that all the observed T_{R_2} are within 0·04 sec. of those calculated, except when $I=0{\cdot}5$ sec. The fit at this point—and indeed at a number of others—would, however, appear to be a good deal closer than is indicated by Fig. 2. Consider, for instance, the case when $I=0{\cdot}5$ sec.: Fe_1 is shown as capturing the central mechanisms thus causing T_{O_2} to be delayed. It will be seen, however, that almost any deviation of T_{R_1} and T_{M_1} from that shown will reduce the delay of T_{O_2}. If $T_{R_1}+T_{M_1}$ is shorter, T_{Fe_1} will be over sooner. If $T_{R_1}+T_{M_1}$ is longer, S_2 will arrive before the end of M_1 and will therefore be in a position to capture the central mechanisms before Fe_1.

Take again the case where $I=0{\cdot}2$ sec. T_{O_2}, as shown in Fig. 2, is slightly delayed because T_{O_1} has not finished when S_2 arrives. Any lengthening of T_{R_1} will increase this delay. On the other hand, should T_{R_1} be reduced to less than 0·2 sec., there will be a substantial likelihood that Fb_1 will capture the central mechanisms and cause T_{O_2} to be delayed during T_{Fb_1}. Thus, any substantial deviation of T_{R_1} from the average value of 0·28 sec. will lengthen T_{R_2}.[1]

[1] On this view we should expect that when I = the mean T_{R_1} of a number of readings, the observed T_{R_2} would have a wider scatter than when I is longer or shorter. The reason is that some T_{R_2} would be delayed by T_{Fb_1} and some would not. Evidence of this wider scatter is contained in Vince's published results. Mr A. Leonard has called my attention to the same result in an experiment by Sándor (11) using a technique very different from that of Vince.

Similar considerations apply when $I = 0{\cdot}3$, $0{\cdot}6$ or $0{\cdot}7$ sec. The expected differences c the observed from the calculated values of T_{R_2} are indicated by + and − signs on th right of Fig. 2. It will be seen that all the deviations are in the expected directions.



I (sec.)	Observed mean T_{R_1} (sec.)	Observed mean T_{R_2} (sec.)	Calculated mean T_{R_2} (sec.)	Discrepancy between observed and calculated T_{R_2} (sec.)	Expected discrepancy
0·05	0·29	0·51	0·53	+ 0·02	0
0·1	0·22	0·43	0·41	− 0·02	0
0·2	0·28	0·41	0·37	− 0·04	−
0·3	0·29	0·40	0·43	+ 0·03	+
0·4	0·28	0·34	0·32	− 0·02	0
0·5	0·26	0·32	0·42	+ 0·10	+
0·6	0·31	0·34	0·37	+ 0·03	+
0·7	0·32	0·33	0·29	− 0·04	−

Fig. 2. Lengthening of reaction time to second stimuli arriving at various intervals after first (13). ■ = S; —— = T_O; ⁓⁓ = T_M; − − − = T_F (shown only when likely to affect T_{R_2}); • • • • = time 'held in store

Similar results were obtained in a second experiment by Vince (16), where the san technique was used, but the steps to be tracked were not all of the same extent. Th observed and calculated T_{R_2} are set out in Table 1. It will be seen that the fit is tolerab

Table 1. *Observed and calculated results of Vince's*(16) *experiment on tracking a step course*

(All times are in decimals of a second)

I	Observed mean T_{R_1}	Observed mean T_{M_1}	Observed mean T_{R_2}	Calculated mean T_{R_2}	Discrepancy between observed and calculated T_{R_2}	Expected discrepancy
0·05	0·24	0·22	0·43	0·44	+0·01	0
0·1	0·24	0·24	0·40	0·39	−0·01	0
0·2	0·25	0·25	0·34	0·30	−0·04	−
0·3	0·25	0·24	0·30	0·35	+0·05	+
0·4	0·25	0·30	0·30	0·25	−0·05	−
0·5	0·28	0·29	0·31	0·25	−0·06	−
0·6	0·25	0·28	0·29	0·33	+0·04	+
0·7	0·25	0·28	0·28	0·25	−0·03	−

T_{O_2} assumed = 0·25 sec. = observed mean T_R for first and single stimuli.
T_{Fb} and T_{Fe} assumed = 0·15 sec. each.

good, no discrepancy between the observed and calculated values being greater than ±0·06 sec. In all cases from $I = 0{\cdot}2$ sec. onwards, the discrepancies are in the directions that variations in T_{R_1} or T_{M_1} would lead us to expect.

(b) Telford(12)

The subject's task in this experiment was to press a key in response to a series of auditory signals arriving at intervals of 0·5, 1, 2 and 4 sec. in irregular order. The mean observed T_R were respectively 0·335, 0·241, 0·245 and 0·276 sec. We cannot represent the results of this experiment with the same confidence that we can Vince's, because we do not know T_M—i.e. the time taken to press and release the key. Some measurements by Dr N. T. Welford and the writer of individual times to press and release a key, and some measurements obtained by Vince(15) of free tapping on a key, suggest that the average T_M is unlikely to have been less than 0·15 sec. If we take

T_{R_1} = 0·254 sec. = the mean observed T_R when $I = 1$, 2 and 4 sec.;
T_{O_2} = 0·241 sec. = the lowest of the observed mean T_R;
T_{M_1} = 0·15 sec.;
$T_{Fb_1} = T_{Fe_1}$ = 0·15 sec.;

the *shortest* mean T_{R_2} that could be expected when $I = 0{\cdot}5$ sec. would be 0·295 sec. as shown in Fig. 3. This is 0·04 sec. less than the figure observed by Telford, but does indicate some delay, and in any case assumes something near a minimum for most of the time components involved.

A possible reason for the increase of T_{R_2} when $I = 4$ sec. is given in the following section.

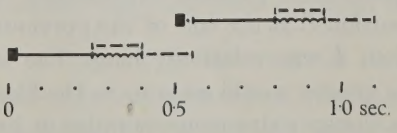


Fig. 3. Lengthening of reaction time to a second stimulus arriving 0·5 sec. after a first(12). Symbols as in Fig. 2.

(ii) *Grouping of stimuli*

One important result of the two experiments by Vince discussed in the preceding section has not yet been mentioned. It is that at the shorter intervals M_1 was sometimes omitted or curtailed before it had come to completion. We shall deal with omissions in this section and with the curtailments in the next.

In Vince's (13) first experiment, M_1 was omitted in 21 % of cases when $I = 0{\cdot}05$ sec., and not at all when I was longer. In the second experiment (16) omission of M_1 occurred in about 20 % of the cases when $I = 0{\cdot}05$ sec., in about 8 % when $I = 0{\cdot}1$ sec., and very rarely when I was longer. A roughly similar result was obtained in another experiment, reported in Vince's (16) paper, in which the subjects were presented with stimuli of two kinds to which ungraded movements in opposite directions had to be made. They were instructed that if a second stimulus to move in the opposite direction to that indicated by the first arrived shortly after the first, they were to omit the first response, but that in any case they were to respond to the stimuli as quickly as possible.

The finding in these experiments that the response to a stimulus can occasionally be inhibited by a second stimulus following shortly after, poses the problem of how this is done if there is refractoriness during T_{O_1}. Three possible ways in which this could occur would seem to deserve consideration:

(*a*) Experiments by Vince (15) have shown clearly that groups of stimuli presented simultaneously can be responded to as single units, and it has been suggested that there is a brief period which may sometimes be as long as 0·05 or 0·1 sec. following the receipt of a stimulus during which a second stimulus can be grouped with the first and the data from both be passed to the central mechanisms together.

(*b*) It is reasonable to suppose that even though a subject is concentrating upon the display, his attention and, in consequence, his central mechanisms will occasionally be taken up by irrelevant stimuli. If a stimulus to which he is supposed to react arrives while his attention is thus diverted, there will be a delay before T_O can begin, as shown in Fig. 4. A second stimulus arriving during this delay would seem likely to be apprehended with the first so that both would be reacted to together, as shown in Fig. 5.

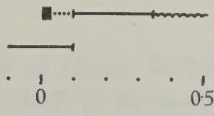


Fig. 4. Delay of single response due to interference by 'irrelevant' central activity. Symbols as in Fig. 2.

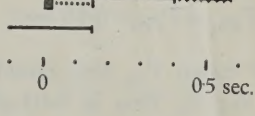


Fig. 5. Grouping of two stimuli because of interference by 'irrelevant' central activity. Symbols as in Fig. 2.

The pattern in Fig. 4 provides a possible explanation of the longer mean T_{R_2} in Telford's (12) experiment (mentioned at the end of the previous section) and in the similar experiment by Hick (6), when I was relatively long. The longer the 'blank' interval between the two stimuli, the greater would seem to be the likelihood of some deflexion of the subject's attention by a chance extraneous stimulus or by wondering when the next stimulus is going to arrive. Any such deflexion would stand a chance of delaying T_{O_2} and so raising the mean T_{R_2}.

(*c*) Vince (16) has suggested that, at least in the third of the experiments so far mentioned, sometimes the subjects may have waited a short time after the arrival of a stimulus and made their reaction conditional upon no other stimulus having arrived during the waiting time. If a second stimulus had arrived by the end of the waiting time the first response would be cancelled and a substitute response made.

The process postulated in (*c*) is well illustrated in an experiment by Hick (4), in which the subjects were given a series of stimuli at irregular intervals and definitely told that when one appeared they should wait for a short time to see whether a second appeared, and if it did, make a double response as quickly as possible. The results suggested that the subjects were waiting an average of about 0·35 sec. before committing themselves to a response. It had previously been ascertained that the reaction time to single stimuli at regular intervals was a little over 0·2 sec. On the basis of these times, we may postulate the scheme set out in Fig. 6. The subjects are shown as waiting for 0·35 sec., and then taking 0·2 sec. to make a decision to respond which is conditional upon no other stimulus having occurred during the interval. This gives a T_R for single stimuli of 0·55 sec., which is close to the observed value of 0·533 sec. If a second stimulus has arrived during the waiting period, the 0·2 sec. is spent making the decision *not* to respond to S_1 and a further 0·2 sec. is spent on the T_O leading to the double response. The double response thus begins at a constant interval of 0·75 sec. after the arrival of S_1, irrespective of the time at which S_2 has appeared. This is close to the value observed: Hick's results show a constant interval of about 0·78 sec.

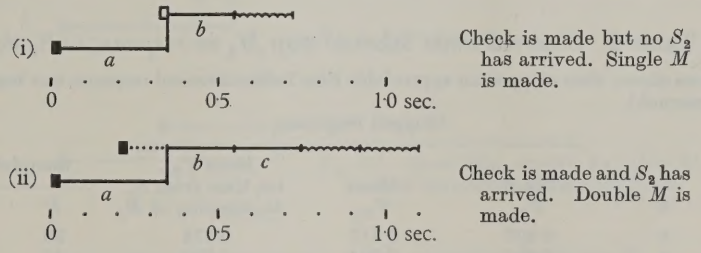


Fig. 6. Voluntary grouping by delay in responding to first stimulus(4). □ = S absent. Other symbols as in Fig. 2. a = delay period (in the nature of a T_O); $b = T_O$ to decide whether or not to make single M; $c = T_O$ to decide to make double M.

This experiment raises a subsidiary point of some importance. The results suggest that the waiting period was, as represented in Fig. 6, in the nature of a T_O. The period was, however, considerably longer than the T_R observed in the experiment, and its duration was presumably under voluntary control. The presence of refractoriness during this period raises the question of whether in some situations complex central processes may produce refractory periods which are very much longer than those observed in the experiments surveyed here. Observation of everyday behaviour seems to provide many examples, but, so far as the writer is aware, none has been subjected to experimental study.

(iii) *Stopping a response*

An experiment by Bates (6) has indicated that once a decision to respond has been taken the response cannot be immediately revoked, and an experiment by Poulton (10) has illustrated that a movement in progress cannot be stopped immediately by a signal to do so. The data recorded in these experiments are, unfortunately, not sufficiently complete

for them to show more than the well-known fact that a reaction time must elapse before a signal to stop a response becomes effective. Poulton's results suggest, however, that refractoriness may have delayed the stopping. The task was to make a tracing movement over a zig-zag course which, upon occasion, had to be either unexpectedly stopped at one of the intermediate angles or extended beyond the expected end-point. Poulton reports that when the signal to do either of these occurred at the point at which the subject was to stop or extend his movement, there was usually an appreciable delay before it took effect. This delay could often be eliminated if the signal was given one angle (about 0·3 sec.) ahead, and could always be eliminated if the signal was given two angles (about 0·6 sec.) ahead. The delays in both stopping and extending were substantially longer than the reaction time for starting the movement. The signal to start was given after a warning, and Poulton interprets his results as indicating that the delays in stopping and extending were due to lack of warning. But it is clear that his results are equally well accounted for by assuming that at each angle and at the end of the movement the subject's attention and central mechanisms were engaged by the sensory data which would be essential for keeping the movement accurate.

Poulton also reports that the delay in stopping was substantially less than that in extending the movement. It is difficult to see from the account of the experiment what this stopping time means, but the result is perhaps plausible in that stopping meant only the cancellation of a movement, whereas extending required both the initiation of a movement and some decision about its direction.

Table 2. *Observed times taken to stop M_1 in response to S_2* (16)

(All the responses shown were stopped an appreciable time before a second response was begun. All times are in decimals of a second.)

		Stopped responses			Completed responses	
I	N	Mean T_{R_1}	Mean T_{M_1}	Mean T_{Re}, i.e. time from S_2 to stopping of M_1	N	Mean T_{R_1}
0·05	3	0·307	0·317	0·574	30	0·224
0·1	9	0·246	0·204	0·350	45	0·219
0·2	9	0·312	0·260	0·372	69	0·226
0·3	4	0·367	0·230	0·297	84	0·229
0·4	2	0·275	0·230	0·105	90	0·238
0·5	1	0·810	0·170	0·480	97	0·260

A clearer indication of the time taken to stop a response is given in the results of Vince's (16) experiment on tracking step functions mentioned in § (i). In dealing with some of the pairs of stimuli the subjects appeared to have curtailed M_1 an appreciable time before beginning M_2. The numbers are small but such as they are they suggest that T_{O_c} (i.e. the time to organize data from a *cancelling* stimulus) was little if any less than T_R for *making* a single movement, and that it was subject to delay due to refractoriness during T_{R_1}.

Vince noted that these stopped movements tended to be preceded by long T_{R_1}. It will be seen from Table 2 that, in fact, the mean T_{R_1} was always longer than I except when $I = 0{\cdot}4$ sec., and the scatter diagrams given by Vince show that this was also true of nearly, if not quite, all the individual readings. In terms of the postulates we have proposed, this means that, except when $I = 0{\cdot}4$ sec., T_{O_c} could begin when M_1 began and T_{M_1} can be taken as a measure of T_{O_c} as shown in Fig. 7*a*. These times are well within the

range of the T_{R_1} recorded in this experiment. For the two stopped responses when $I = 0{\cdot}4$ sec., T_{O_c} cannot have been more than 0·105 sec., which is rather short, but still just within the range of the observed T_{R_1}. The time relationships postulated for these two responses are shown in Fig. 7*b*.

The reason why these stopped movements hardly ever occurred when T_{R_1} was less than I is suggested in Fig. 7*c*, which shows T_{O_c} being delayed by T_{Fb_1} until it is too late to take effect.

The association of long T_{R_1} with curtailment of M_1 cannot be explained in this way when I is as short as 0·05 or 0·1 sec., because the T_{R_1} required on this hypothesis would be too short. There would seem, however, to be two ways in which this association might arise:

(*a*) It is possible that there was some tendency for slower movement to go with longer reaction times, in which case there would be more likelihood of the movement following a long reaction time being stopped as a result of a second stimulus. An association of this kind has been shown in another experiment by Vince (14).

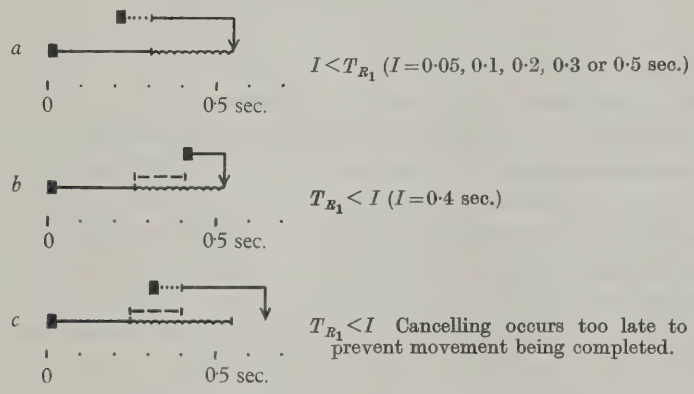


Fig. 7. Stopping a movement (16). ↓ =stopping of previous M. Other sympols as in Fig. 2.

(*b*) It is possible that some of these curtailed M_1 were made in cases where the two stimuli were grouped. The instructions given to the subject had the slight ambiguity that if such grouping occurred he might be uncertain whether he should omit the first response entirely or reduce it to a kind of token response. We have already seen that grouping of this kind may be associated with a lengthening of T_{R_1}.

Finally, we may note under the heading of this section a special case of stopping reported by Poulton (10). In this experiment, two stimuli were presented after a warning signal. The subject was prepared for them both, but was instructed to begin his first response on receipt of the first stimulus. Three values of I were used: 0·1, 0·2 and 0·4 sec., but only one of these was used in any one session so the subject not only knew what stimuli to expect but also when to expect them. The findings were that T_{R_2} was substantially longer than T_{R_1}, when $I = 0{\cdot}1$ sec., slightly longer when $I = 0{\cdot}2$ sec., and the same as T_{R_1} when $I = 0{\cdot}4$ sec.

This experiment would seem not to be strictly comparable with those of Vince because, when both S_1 and S_2 are expected, S_1 is likely to become the stimulus for a double response

to the making of which S_2 is more or less irrelevant. That something of this kind was happening is suggested by Poulton's finding that in four out of the sixteen occasional 'catches' when $I = 0{\cdot}2$ sec. and S_2 was replaced by a different S to which the subject had been instructed not to respond, M_2 was nevertheless made. It is again, unfortunately, difficult to represent this experiment in diagrammatic form because we have no knowledge of T_{M_1}. We cannot, therefore, be certain whether the lengthening of T_{R_2} when $I = 0{\cdot}1$ and $0{\cdot}2$ sec. was due to some refractory period phenomenon or whether it was merely the result of the time taken to carry out M_1. What we must try to do is to explain how M_2 was stopped in the twelve out of sixteen cases when the substitute second stimulus was given with $I = 0{\cdot}2$ sec. T_{R_1} averaged a little over $0{\cdot}2$ sec. so that, on the average, S_2 arrived just before the beginning of M_1. It is likely, therefore, that the substitute second stimulus was able to capture the central mechanisms and become effective whenever T_{R_1} was of average length or longer. When, however, T_{R_1} was less than $0{\cdot}2$ sec. so that M_1 had begun before the substitute second stimulus appeared, T_{Fb_1} would be likely to delay the stopping of M_2 until it was too late.

(iv) *Amending a response*

(*a*) *Amending stimulus independent of first response*

Direct published evidence on this point is at present lacking except for Poulton's[10] experiment which we have already mentioned. In the terms we have previously used, we should expect that an amending stimulus would take effect at a time which was dependent upon when it arrived in relation to the time of onset of the original response. If S_a (i.e. a stimulus indicating that a previous response should be *amended*) arrives before M_1 has begun we may expect T_{O_a} to begin when M_1 begins. Thus M_a will begin $T_{R_1} + T_{O_a}$ after S_1, irrespective of the exact time when S_a appears. This relationship might not hold completely when I was very short owing to the likelihood of the two stimuli being grouped. Where this occurs, M_a should replace M_1 completely. If S_a arrives after the beginning of M_1 we should expect M_a to be delayed by any T_F that might occur. If, for instance, a T_{Fb} occurs, we should expect M_a to begin $T_{Fb} + T_{O_a}$ after the beginning of M_1. The time relationships expected are shown in Fig. 8.

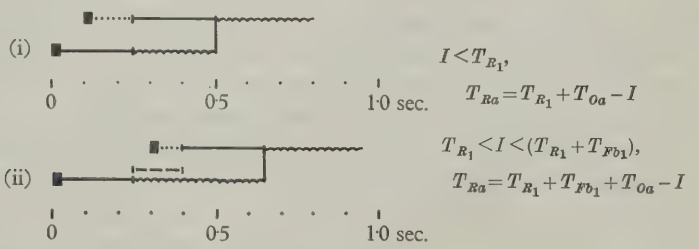


Fig. 8. Amending a response. Amending stimulus independent of first response. Symbols as in Fig. 2.

(*b*) *Amending stimulus dependent upon the beginning of the first response*

Under this head come two experiments in which the subject had either to press a lever or turn a knob against spring tension to move a pointer on to a target line.

Vince (14). The subject had to move a lever against a light spring tension which was occasionally and without warning increased to a point at which the correct pressure for the light tension would produce no movement. Vince found that T_R was, with a light tension, 0·19 sec. on the average, and with the heavier spring tension 0·35 sec. Subtracting the first of these from the second, she arrived at the figure of 0·16 sec. as the average T_{Ra}. This figure would seem by no means impossibly short, provided it could be argued that since S_a was itself a feedback from the beginning of the response no delay due to T_{Fb} occurred.

Hick (5) has, however, criticized this experiment on the ground that the inertia of the subject's arm might cause some movement to be made in response to the original stimulus in spite of the increased spring tension. Certainly, if the published examples of amended responses in Vince's paper are typical it would appear that this did in fact happen, and that M_a began about 0·5 sec. after S_1—i.e. a little over 0·3 sec. after the beginning of the first abortive response. This is the figure that Hick obtained in a closely similar experiment of his own, where inertia of the arm was excluded by confining the movement to one of the wrist or finger. The remarkable point about Hick's finding is that if the scheme we have proposed is correct, and if T_{Oa} had to wait until a T_{Fb}, of 0·15 sec. was over before it could begin, T_{Oa} works out at about the same figure as that found by Vince. It should be remembered, however, that in a case like this where S_a coincides with the beginning of the first response, we should expect T_{Ra} to be variable. It is possible, for instance, that Hick's average figure of 0·304 sec. was made up of some cases where S_a with a T_{Oa} of say 0·19 sec. (i.e. the same as T_{R_1}) managed to capture the central mechanisms before any Fb_1, and of some cases where T_{Oa} had to wait until a T_{Fb_1} was over, thus producing a T_{Ra} of 0·34 sec. (0·19 + 0·15). We cannot distinguish between these two possibilities on the basis of the published data because no scatter of readings is given, but it is clear that if the first of these alternatives is correct, the distribution of T_{Ra} should be that of a single population, whereas if the second is correct the distribution of T_{Ra} should be a combination of two populations.

(c) *Amending stimulus dependent upon the results of the first response*

Under this head we may consider three experiments reported by Vince (14). The first two of these used the technique of pressing a lever against a spring tension and the third used an apparatus similar to that previously used for tracking step functions. The purpose of these experiments was to study the accuracy of movements in relation to their duration. Vince mentions the fact noted by Woodworth (18) that short movements are formed as a whole before they are started and then 'fired off', so to speak, without any sensory control during their course. Slower movements may, however, be controlled and amended by observation of results.

It is obvious that a visual stimulus for the amendment of a movement cannot occur until after the movement has started, so that T_{Oa} could not begin until after any T_{Fb} had finished. M_a could, therefore, not begin until $T_{Fb} + T_{Oa}$ after the beginning of M_1. Taking our previous value of 0·15 sec. for T_{Fb} and Vince's 0·25 sec. for T_{Oa}, M_a could not begin before an average of 0·4 sec. after the beginning of M_1. Some further time would then have to elapse before M_a was complete.

This fits well with Vince's results. Her data showed clearly that movements having durations of 0·6 sec. and upwards were all of about the same average accuracy, whereas

faster movements showed rapid deterioration of accuracy as duration decreased. In the first of the three experiments, the error in movements of 0·4 sec. was, under normal conditions of working in which the subjects could see what they were doing, as great as the error made in slower movements when they shut their eyes as soon as the movement was begun and were thus unable to exercise visual control.

(v) *Continuous performance at tracking tasks, etc.*

The relations discussed under iv(*c*) are relevant to corrections in continuous tracking tasks. The stimulus for correction must follow the beginning of the previous correcting movement and the T_{O_a} cannot, therefore, begin until any T_{Fb} of the previous movement is over. It has several times been indicated in tracking tasks that a time lag of about 0·3 sec. occurs at points where there are sudden changes of direction in the track. If we take this as the T_{O_a} and give the T_{Fb} its previous value of 0·15 sec., we obtain a maximum average rate of correcting of one correction per 0·45 sec., as shown in Fig. 9. This is a little less than the figure of one correction per 0·5 sec., which has been quoted in the past, but would seem to be in fact a better fit to some at least of the data (see Vince (13), p. 150, fig. 1; Welford *et al.* (17), exp. 4).

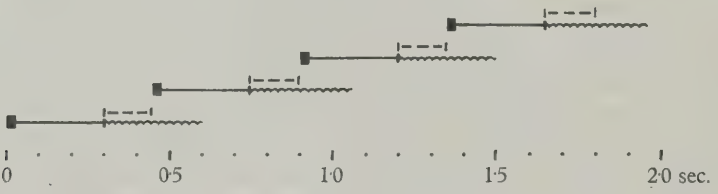


Fig. 9. Successive movements where each partly depends on the one previous. Symbols as in Fig. 2.

The time relations of Fig. 9 offer two important further suggestions:

(*a*) All the experiments considered so far have given the subjects a task in which their performance has been paced by the display. It would seem, however, that essentially the scheme proposed ought to apply also to performances which are carried out at a speed determined by the subject himself.

(*b*) The rate of responding shown in Fig. 9 is determined by the two time components T_O and T_{Fb}. It would appear that if either of these could be reduced or eliminated, a substantially higher rate could be attained.

We may consider these two points in turn.

(*a*) *Unpaced tasks*

The writer's attention was called to the possibility of linking paced and unpaced tasks in a common theoretical framework by some experiments by Leonard (7), and by a striking correspondence between two experiments, one by Brown and one of his own, in which the performances of different age groups were being compared (see Welford *et al.* (17)). The writer's own experiment used a high-speed tracking task in which the subject had to keep a pointer, which he could move from side to side by means of a handle, in line with a 'target' pointer moved by an irregularly shaped cam. It was found that the performance of a group of subjects in their late 'teens and twenties—the same age as those studied by Vince and others—showed clear signs of breakdown when the average swing of the

'target' pointer took less than about 0·5 sec. A similar breakdown among a group of older subjects, mostly in their fifties and sixties, was observed when the average swing of the 'target' pointer took less than about 0·7 sec. Brown's experiment was a tracing task. The subjects had to trace with a stylus at high speed over metal figures 1, 2, 3, 4, 5, 6, 7, 8, 9, 0, inlaid in a metal plate. A calculation of the number of discrete movements required in this task was made, taking the beginning and the end of each figure and each point at which a sharp change of direction had to be made (e.g. at the bottom of 2 or the middle of 3) as marking the beginning of a new movement. The average times per movement calculated on this basis were found to be: for subjects in their twenties about 0·45 sec., and for subjects in their fifties and sixties about 0·65 sec. We were unfortunately not able to record the performances of subjects in sufficient detail to say whether this correspondence

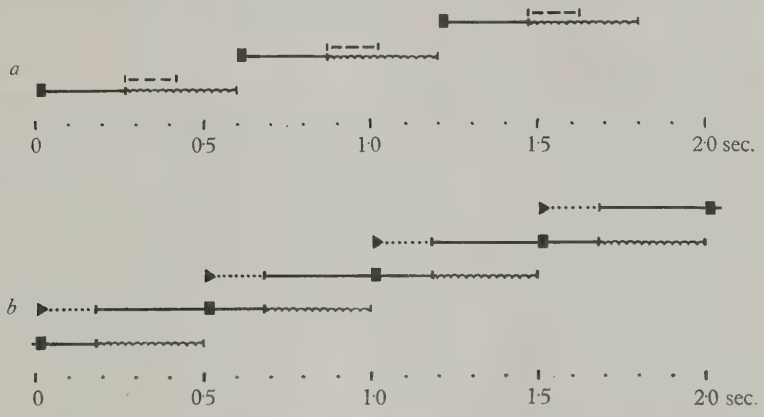


Fig. 10. Successive responses with and without fore-knowledge(7). *a*, Without fore-knowledge. Stimulus appears at completion of previous response. *b*, With fore-knowledge. Each stimulus accompanied by an advance signal of the next. ▶ =advance signals. Other symbols as in Fig. 2.

between the two experiments was more than fortuitous, but such correspondence would appear to be very plausible. Each movement in Brown's experiment had to be made in relation to the point which had been reached by the preceding movement. Thus the stimulus for a movement could not occur until after the preceding movement had begun, and because of this the central mechanisms were liable to capture by Fb. It would seem, therefore, that the minimum time in which each discrete movement was likely to have been made was, as in the tracking case considered above, $T_O + T_{Fb}$. Some subjects, admittedly, did attain considerably higher speeds than this. This they might do if they were able to group some of the movements together, e.g. if they were able to trace each figure as a single whole. There is, however, an alternative way in which it is conceivable that higher rates could be attained, which will be considered later.

Leonard's(7) experiment could be presented in two ways. In the first, the task was to move a lever, working in a triangular guide, to the corner of the triangle where a light appeared. On reaching this corner the light would go out and another would appear in a different corner, and so on. The task in the second condition was the same, except that with each light a second light appeared which gave the subject fore-knowledge of which light would come on next in the series.

The time relationships for the two conditions, according to the scheme proposed, are shown in Fig. 10. The T_M shown are the averages observed by Leonard for moving from

each corner to the next. The gap between each T_M and the next represents the observed average time spent in each corner and can be regarded as a T_R for the data from the S which has appeared on the completion of the preceding M. Comparing Fig. 10*a* with *b*, it will be seen that T_R is shorter in the second condition than in the first, presumably due to the fore-knowledge given by the secondary light enabling the decision about the direction of each M to be taken during the preceding T_M. The giving of fore-knowledge in the second condition means that the next S can capture the central mechanisms at the beginning of each M and thus enable T_O to replace the T_{Fb} shown for the first condition.

It will be seen that in the second condition all attention to F is shown as omitted. This was strikingly confirmed by the introspective reports of the subjects, many of whom reported surprise that in this condition, in contrast with the first, they had very little awareness of what they were doing—the performance seemed automatic and without conscious control.

(*b*) *The possibility of higher rates of response*

If the time relationships shown in Fig. 9 are correct, there seems to be no reason why, if T_O could be reduced to about 0·2 sec. or if T_{Fb} could be eliminated (as shown in Fig. 11), responses could not occur at a rate of about 3 per sec. Indeed, if both could be accomplished simultaneously, responses should occur at about 5 per sec. In this connexion it is interesting to note that in the writer's tracking experiment, where there was evidence that T_O was about 0·3 sec., the subjects maintained the correct number of swings of their pointer even when the 'target' pointer was swinging at an average rate of about three times per second, although their movements were then very inaccurate. In Brown's experiment also, a few exceptionally quick subjects managed to cover the figures at a speed which would indicate that they were making about three movements per second, although again with a substantial loss of accuracy.

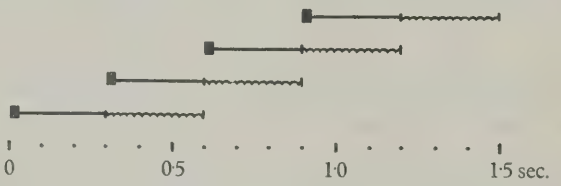


Fig. 11. Successive responses at maximum rate: T_F eliminated. Symbols as in Fig. 2.

A similar result was obtained by Vince (15) in an experiment in which a series of dots on a moving band was shown through a slit in a screen. Dots followed each other at regular intervals, except for an occasional omission, from 1 sec. per dot through intervals of 0·5, 0·3, 0·25, 0·17 and 0·14 to a maximum rate of 0·1 sec. per dot, then back again through the intervals to 1 sec. The results indicated that responses were very accurate at the speeds of 1 and 0·5 sec. per dot, but that accuracy fell off between this speed and the speed of 0·3 sec. At rates higher than this, responses appeared to become chaotic, although responses up to about 3 per sec. were made.

The suppression of all attention to sense data fed back from the response would make little difference to accuracy in a case such as Leonard's where the response movement was ungraded, and knowledge that the required movement had been made correctly was given by the arrival of the succeeding stimulus.

If either of these conditions is unfulfilled, however, accuracy would seem likely to suffer. Let us consider the three experiments in turn.

If performance in the writer's tracking experiment at high speeds can be represented by Fig. 11, it is clear that the stimulus for each movement is entirely independent of the results of the previous movement. In so far as it gives any knowledge of results it is knowledge of the results of the last movement but one.

A subject in Brown's figure-tracing experiment could conceivably achieve the pattern represented by Fig. 11 if he could at the start of each movement immediately transfer his attention to the next movement required and so prevent *Fb* capturing his central mechanisms. But if he did this, he would have no chance, when he made an error in one movement, of correcting it in the next. His first chance of correcting would be in the next but one.

It is doubtful whether the pattern of times in Fig. 11 applied to Vince's experiment. We do not know what T_O was, but by analogy with other reaction time experiments which call for a simple ungraded response to visual stimuli it is likely to have been about 0·2 sec. If it was, and if the time relations were similar in kind to those of Fig. 11, we should have expected a response rate of 5 per sec. instead of the 3 per sec. which was observed. This experiment differs from the others, however, in the important respect that the stimuli gave no information at all about the occurrence of the preceding response movements—the arrival of the stimuli was entirely uninfluenced by the subject's responses. If, therefore, he had attained a response rate of 5 per sec. by omitting all attention to sense data coming from the response movements, he would have had no detailed confirmation whatever of the adequacy of his performance. A subject might be able to perform under these conditions, but he would be unlikely to do so with any confidence, and it would seem probable that some attention to the response would be compulsive. Now it is clear that with the times postulated—$T_O = 0{\cdot}2$ sec. and $T_{Fb} = T_{Fe} = 0{\cdot}15$ sec.—there is time, if stimuli arrive at intervals of 0·5 sec., for all three components to occur with every reaction. If stimuli arrive at a higher rate than this, something must drop out. It would seem plausible to suppose that what happened when I was less than 0·5 sec. was that T_{Fe} dropped out, leaving a pattern of time relations similar in kind to that of Fig. 9. With the times postulated this would mean that responses would be made at an average rate of one per 0·35 sec. (0·2 + 0·15) which is very close to that observed.

(c) *Further relations with skilled performance*

It is conceivable that with intensive practice the movements required in Brown's figure-tracing task could have been learnt to a sufficient level of accuracy for them to be carried out without detailed knowledge of results and thus for performance to be accurate with the time relations of Fig. 11. Such a performance would be very fast, but its maintenance would be dependent upon three conditions being fulfilled:

(i) The subject's attention would have to be continually concentrated on the display one step ahead of the movement.

(ii) Accuracy would be very much dependent upon the maintenance of a precise learned effector performance.

(iii) The subject would have to work with little, if any, knowledge of the details of his performance.

It would seem possible that certain high-speed skills, such as those in fast ball games

and some industrial operations, are performed in this way. The characteristics of these skills are strikingly similar to the conditions outlined above, thus:

(*a*) There is stress on attention to the display—'keeping one's eye on the ball'—and the performance has a clearly anticipatory character.

(*b*) Such skills are usually acquired and maintained at a high level only by long and continual practice.

(*c*) Those who perform them best are often almost completely unaware of the details of what they do (see Bartlett (1)).

It seems that the comparison of these skills with the type of performance indicated by Fig. 11 can be pressed further. It is known that when performance breaks down, e.g. in so-called 'fatigue', it often becomes slower and at the same time the subject has increased awareness of what he is doing (see Davis (3)). Fig. 11 suggests that any impairment of the subject's ability to keep his attention one step ahead of his effector performance, or any decrease of effector accuracy, would necessitate a reversion to a pattern like that of Fig. 9, or in severe cases to that of Fig. 12. If this occurred, performance would certainly

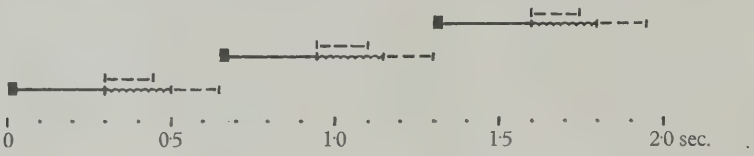


Fig. 12. Successive responses when both T_{Fb} and T_{Fe} occur. Symbols as in Fig. 2.

slow down, because there would be opportunity for 'feedback' data to capture the subject's attention. It should also be noted that in any task demanding continuous corrections—e.g. a tracking task—the slower speed of correction would mean that the size of misalinement which would occur before correcting action was taken would increase (see North (9), p. 23).

III. Summary

Experiments by several authors have shown that when two stimuli requiring hand action are presented close together, the reaction time is frequently longer to the second than it is to the first. Some previous interpretations are surveyed, and the following theory is put forward:

1. The time required by the central mechanisms to deal with the information provided by a stimulus and to initiate a response to it is, on the average, the same whether the stimulus comes closely after another or not. We may call this time the *organizing time*. Normally it will be the same as the reaction time for a single response.

2. The central mechanisms are liable to become engaged by stimuli fed back from the response, particularly from those parts of it where there is rapid acceleration or deceleration of movement, or where some definite sound or visual or tactile change is produced. The perception of any such feedback data will require central organizing time.

3. No two central organizing times can overlap, so that information from a stimulus arriving while information from a preceding stimulus is being dealt with has to be 'held in store' until the central mechanisms are free.

A survey of the evidence indicates that this theory, with certain assumptions about the duration of central processes, can account for the time relations observed not only in serial reaction time experiments of various kinds, but in tracking and other continuous

performance tasks. Special cases where apparent exceptions occur seem capable of being accounted for by reasonable additional hypotheses.

My sincere thanks are due to Prof. Sir Frederic Bartlett for several valuable suggestions, and to Miss M. A. Vince, Mr A. Leonard and Dr E. C. Poulton for their generous advice and constructive criticism. My especial thanks are due to Dr W. E. Hick from whose published work and discussions of the preliminary drafts have come the beginnings of many of the ideas contained in this paper.

References

(1) Bartlett, Sir Frederic C. (1947). The measurement of human skill. (Oliver Sharpey Lectures to the Royal College of Physicians of London.) *Brit. Med. J.* I, 835, 877; *Occup. Psychol.* 1948, XXII, 31, 83.

(2) Craik, K. J. W. (1948). Theory of the human operator in control systems, II. *Brit. J. Psychol.* XXXVIII, 142.

(3) Davis, D. R. (1948). *Pilot Error.* Air Ministry Publication. A.P. 3139 A. London: H.M.S.O.

(4) Hick, W. E. (1948). Discontinuous functioning of the human operator in pursuit tasks. *Quart. J. Exp. Psychol.* I, 36.

(5) Hick, W. E. (1949). Reaction time for the amendment of a response. *Quart. J. Exp. Psychol.* I, 175.

(6) Hick, W. E. & Bates, J. A. V. (1950). *The Human Operator of Control Mechanisms.* Ministry of Supply Permanent Records of Research and Development Monograph No. 17.204.

(7) Leonard, A. (1950). Unpublished report at the Psychological Laboratory, Cambridge.

(8) Mowrer, O. H. (1940). Preparatory set (expectancy)—some methods of measurement. *Psychol. Monogr.* LII, no. 233.

(9) North, J. D. (1950). *The Human Transfer Function in Servo Mechanisms.* Ministry of Supply Report No. WRD/6/50.

(10) Poulton, E. C. (1950). Perceptual anticipation and reaction time. *Quart. J. Exp. Psychol.* II, 99.

(11) Sándor, B. (1931). Die Reactionszeiten des Menschen. *Industr. Psychotech.* VIII, 233.

(12) Telford, C. W. (1931). Refractory phase of voluntary and associative responses. *J. Exp. Psychol.* XIV, 1.

(13) Vince, Margaret A. (1948*a*). The intermittency of control movements and the psychological refractory period. *Brit. J. Psychol.* XXXVIII, 149.

(14) Vince, Margaret A. (1948*b*). Corrective movements in a pursuit task. *Quart. J. Psychol.* I, 85.

(15) Vince, Margaret A. (1949). Rapid response sequences and the psychological refractory period. *Brit. J. Psychol.* XL, 23.

(16) Vince, Margaret A. (1950). Some exceptions to the psychological refractory period in unskilled manual responses. Medical Research Council Applied Psychology Unit Report No. 124/50.

(17) Welford, A. T. *et al.* (1951). *Skill and Age: an Experimental Approach.* Oxford University Press for the Nuffield Foundation.

(18) Woodworth, R. S. (1899). The accuracy of voluntary movement. *Psychol. Rev. Monogr. Suppl.* III, 13.

(*Manuscript received* 29 *May* 1951)

MEMORY AND INTELLIGENCE[1]

By J. G. INGHAM

Psychological Department, Institute of Psychiatry, London

I. *The problem* (p. 20). II. *The investigation* (pp. 20–22): (i) *General plan* (p. 20); (ii) *Practical details* (p. 21). III. *The tests* (pp. 22–25): (i) *Intelligence tests* (p. 22); (ii) *Memory tests* (p. 22). IV. *Results* (pp. 22–29). V. *Discussion* (pp. 29–31). VI. *Summary* (p. 31). *References* (p. 32).

I. The problem

A *memory process* may be defined as one in which learning modifies later behaviour or experience. Assuming this modification to be mediated through some physiological change within the organism, three phases of the process may be distinguished. In the phase of *impression* the organism interacts with its environment in some way and is modified. During the *retention* phase, the modification is maintained. Finally, the behaviour or experience of the organism occurs in a modified form and this will be called the *re-emergence* phase.

Can inter-individual differences in the retention phase of all memory processes be entirely accounted for by differences in general intelligence plus specifics, or is it necessary to postulate the existence of another factor influencing retention in all circumstances? An evaluation of fourteen investigations published during the period from 1915 to the present day, revealed conflicting results but little evidence of a broad retention factor (5). Whenever a wide range of memory tests was employed, the results were essentially negative. Owing to the nature of the tests used in these investigations, however, it is impossible to say that they provide evidence against the existence of such a factor. This is because (*a*) the retention intervals were extremely short (nearly all were tests of immediate memory), and (*b*) no attempt was made to ensure that all individuals learned the material to the same level of proficiency. Under these conditions it can be shown that the influence of the retention phase on the final score has been minimized. The dice, therefore, have been loaded against the appearance of significant residuals when the influence of variables other than those operating during the retention phase has been eliminated.

Where tests of delayed memory have been used, variables such as retroactive inhibition and reminiscence, operating in the retention phase, have not been adequately controlled, again tending to reduce the proportion of score variance due to retention. The object of this investigation was to overcome these difficulties and obtain more definite evidence relating to the problem stated above.

II. The investigation

(i) *General plan*

The memory tests used were devised to ensure that the scores would depend as far as possible on the retention phase. This was achieved by controlling the learning situation in such a way that each subject learned to about the same level of proficiency. By this

[1] Based on a thesis approved for the degree of Ph.D in the University of London. The research was initiated at the suggestion of Dr H. J. Eysenck.

means the influence of variations in the learning phase[1] on the final score was minimized, he final score being the amount retained after a constant time. All the tests used in the nvestigation were administered individually.

The retention interval was fixed at 30 min. This was found long enough to allow ppreciable forgetting to take place for most subjects and short enough to be within he control of the investigator during one testing session. Each session was kept easonably short to avoid undue fatigue in the subjects.

It was necessary to keep the subjects occupied during the retention interval, in order o minimize the tendency to review the material learned. As a measure of intelligence vas also required, the most economical procedure was to occupy this time interval with n intelligence test. Each testing session was therefore divided into three phases:

Phase I. Learning all the memory material to a fixed criterion (Learning phase).

Phase II. Time interval, occupied by intelligence tests (Retention interval).

Phase III. Testing the amount retained (Re-Learning phase).

(ii) *Practical details*

Each individual was tested in two sessions, each session being organized according o the above scheme, on two mornings of the same week. Four subjects were tested er week, over a period of about 5 months. The sessions required from 1 to $1\frac{1}{2}$ hr. each, naking a total of about $2\frac{1}{2}$ hr. testing for each subject. Most of the subjects appeared to naintain a reasonable interest in the tests over this period.

During Session 1, memory tests B, C, A and D (see Table 1) were learned consecutively, n the order stated. Under these conditions, one would expect retro-active and pro-active nhibition to result from the interaction of the various learning processes. In order to educe this to a minimum, the material learned during one session was varied as much as ossible. The time of beginning and end of learning was noted, to the nearest minute, or each test. The retention interval was occupied by subtests of the Wechsler Bellevue ntelligence Scale (Form II), namely Comprehension, Digit Span, Similarities, Picture Arrangement and Block Design. As far as possible they were given in the same order or each subject but, as the times required varied from subject to subject, some adjustnent was necessary to ensure that 30 min. intervened between learning and re-learning or each test. By this method it was possible, with few exceptions, to make the retention nterval constant to within ± 2 min. It was not usually necessary to give all the Wechsler tests to occupy the time interval between the end of learning and the beginning f re-learning. The remainder were given either during the re-learning phase in the ntervals between the memory tests (the time was adjusted for each test independently) r after the completion of re-learning.

Session 2 was administered in precisely the same way. The memory tests were given in he order G, H, E and F (see Table 1). The Wechsler tests used for this session were Arithmetic, Vocabulary, Picture Completion, Object Assembly and Digit Symbol.

An analysis of the retention intervals showed that, for the majority of subjects on all ests, the times varied from 28 to 32 min. The largest number to lie outside this range or any one test was 11. It is not likely therefore that variations in the length of the etention interval had an appreciable influence on the scores.

The total number of subjects tested was 80, of whom 77 were N.C.O.'s and O.R.'s

[1] The identification of learning with the impression phase of the memory process is only approximate.

from the 2nd Battalion Coldstream Guards[1] and two were students from a University Extension class in psychology. The other subject was a member of the staff of the Institute of Psychiatry, Psychological Department. With one exception, they were male and ages ranged from 18 to 39. Only two were older than 31.

The average I.Q. of the group, as measured by the Wechsler Bellevue Form II (Full Scale), was 108·8 with a standard deviation of 11·8. A comparison of these figures with those given by Wechsler shows that the group used in this investigation had a higher mean and lower standard deviation than the standardization group for Form I of the same test. Although insufficient evidence is available to assume that Form II is strictly equivalent to Form I, it is standardized to give a mean of 100 and a standard deviation of about 14·7. It does seem highly probable, therefore, that this group is above the average of the standardization group and is more homogeneous, with respect to the abilities measured by this test.

III. The tests

(i) *Intelligence tests*

The choice of intelligence tests for this investigation was based upon the dual purpose of assessing intellectual ability on a range of activities and occupying the time interval between learning and re-learning, to prevent the subjects from reviewing the material. To achieve these objects, it was necessary to have an individual test, covering a reasonably wide range of abilities, and including a number of separate subtests of varying lengths. The Wechsler Bellevue Intelligence Scale was chosen, Form II of the test being used as it was most readily available, in this department, at the time of the investigation. The General Information subtest was omitted, as the questions are unsuitable for English subjects. Alterations were made in some of the Comprehension and Arithmetic questions, though the sense and the actual computations were maintained. In the Vocabulary test, 'dollar' was substituted for 'shilling'. Apart from these alterations, the material, procedure and scoring were exactly as stated in the manual (10).

(ii) *Memory tests*

Bearing in mind the deficiencies observed in previous memory tests, the aim was to devise tests which would meet the following requirements:

(*a*) Scoring should be as objective and simple as possible.

(*b*) The testing procedure should be constant for all subjects.

(*c*) The tasks to be learned should be well within the subject's capabilities, to avoid undue strain.

(*d*) The tests should cover a wide range of material.

(*e*) The retention interval should be long enough to allow appreciable forgetting to take place, yet short enough to allow conditions to be carefully controlled.

(*f*) At least one score should be available to measure the amount retained, with variations due to initial learning eliminated.

(*g*) In addition, it should be possible to assess other phases of the memory process, for example, learning ability.

[1] I am indebted to Col. L. Bootle-Wilbraham, D.S.O., M.C., and to all those who acted as subjects.

Points (*a*) and (*b*) need no elaboration as they are essential requirements of any objective mental test. Point (*c*) is also a desirable characteristic of any testing situation in which a subject's co-operation is required, but it deserves special mention in the case of learning tests. The usual type of laboratory learning experiment is carried out under conditions which are tedious in the extreme for the subject. This is, of course, justified when the subjects are in full sympathy with the aims of the experiment and can be relied upon to give their full co-operation, but the mental tester rarely has such a favourable situation. The problem has been largely overcome in the case of intelligence testing, by making the test material as interesting as possible. In the memory tests used in this investigation, boredom was reduced to a minimum by keeping the tests short and varying the material. The advantage that might have been gained by having a wider range of scoring was thus sacrificed.

The importance of (*d*) can be seen by considering one of the criticisms which have been directed against some of the earlier researches in the field of memory. This is that the memory factor, when found, has covered such a restricted range of tests that it does not warrant interpretation as a broad group factor. If such a criticism is to be avoided, therefore, it is clear that the learning material should be drawn from as wide a field as possible.

Point (*e*) arises from the necessity for the retention scores to cover the range of abilities in the subjects tested. The ideal arrangement would be for the most retentive subject to remember nearly all the material and for the least retentive to forget nearly all, but it would be satisfactory if very few attained either the minimum or the maximum score. The other factor limiting the time interval is that it should be short enough to be within the control of the investigator. As is shown clearly in the literature on the subject of retroactive inhibition, one of the most important conditions governing retention is the amount and sort of activity in the intervening time interval(6). If retention were to be measured over a period of several hours or more, it would be difficult to govern the way in which the subjects passed their time. Some would rehearse the material while others would not. Some would indulge in activities which would interfere quite strongly with the learned responses, while others would rest. In fact, the amount retained would vary from subject to subject for reasons quite extraneous to the true variable which the tests are designed to measure. It was therefore deemed necessary to limit the time interval so that it could be entirely contained within one testing session.

One aim of this research was to devise tests which would measure that part of the memory process which did not involve initial learning. For this purpose (*f*) is an essential requirement. It was also considered desirable, however, to see how far other phases in the memory process involved the same factor or factors and, for this purpose, (*g*) was added.

The test material, which was presented in the form of paired-associates, can be classified under the following headings:

) *Meaningful*

Both stimulus and response items were either:

(i) Verbal...familiar words.
(ii) Pictorial...drawings of objects.

(2) *Meaningless*

The stimulus items were meaningful but the response items were either:

(i) Verbal...nonsense syllables.
(ii) Pictorial...meaningless drawings.

There were two tests in each of the four categories.

It must be admitted that only a restricted range of the meaningful-meaningless dimension has been covered by these tests. Nevertheless, it is believed that the material is sufficiently varied to ensure that if a memory factor seems to be required to explain the correlations, then it must be of some importance. We shall only be justified in referring to it as a 'rote memory' factor, and further research will be required to elucidate the relation between retention under these conditions and retention of meaningful material.

The method of learning was chosen to meet requirements (*a*), (*b*) and (*g*). It was necessary for learning to be carried out under carefully controlled conditions and, as far as possible, to be assessed quantitatively at each stage. The procedure that was chosen to fulfil these conditions most adequately was that of paired-associates, with anticipation and prompting on each repetition. The items were presented visually, each pair on a card 5×3 in. The cards were exposed successively for 2 sec. each, judged by means of a stop-watch, on an exposure stand designed for the purpose. After the presentation of the whole series, the first item of each pair was exposed and the subject was asked to indicate what the second item was. In four of the tests he was asked to recall the second item and in the others he was asked to select the correct one from a list containing all the response items (multiple choice). If he failed to respond correctly, the response item was then exposed. If he gave the correct response, he was told it was correct and the next stimulus item was exposed. This was continued until the learning criterion had been reached. The order of presentation of the items was the same throughout.

There was one exception to the procedure detailed above. It was found that, even restricting the number of pairs to three, test D was too difficult for many of the subjects tested during the preliminary investigations. For this reason, it was decided to modify the learning procedure somewhat. After the first presentation of the first pair, the stimulus item was exposed and the subject asked to draw the correct figure to go with it. If he failed, the response item was exposed once more and the subject was given another opportunity to recall it correctly. This was continued until one correct drawing was made and then the next pair was exposed. The same procedure was followed for each of the three pairs of items, and from then onwards the series was presented in the same way as the other tests, giving one opportunity for recall of each item on each repetition. In this way the test was brought within the capabilities of most of the subjects, but some still failed to complete it. It had, therefore, to be excluded from the final analysis except for the immediate memory score which was available for everybody. The re-learning phase, for all tests, was administered in precisely the same way as the learning phase, except that the subject did not see the response items before his first attempted response.

The learning criterion used was that of one correct response to each item. In order to avoid over-learning of certain items, which inevitably occurs if learning is continued until the list as a whole is repeated correctly, each item recalled was omitted from later

epetitions of the material. This method, which was used by Woodworth (11), does not ntirely eliminate variations in the amount of initial learning, as it introduces a bias n favour of the faster learners. If one considers the extreme cases, for example, it is lear that the subject who makes one correct response per trial has learned the material ess adequately than the one who learns all the items on the first repetition. Although he two extremes rarely arise, the same relationship holds for the intermediate cases. For this reason, it is impossible to say that the amount retained is entirely uninfluenced by the initial learning level, though the effect of the latter is reduced to a minimum.

The following methods of scoring were used to assess the various stages of the memory process:

1) *Learning score*

This was intended to give a measure of the subjects' learning abilities and was the number of repetitions required to learn each item, summed over all the items.

2) *Retained items*

The retained items score was a measure of the amount retained after 30 min. It consisted of the number of items recalled or recognized correctly, on the first attempt of the re-learning phase.

3) *Saving score*

This was an alternative assessment of the amount retained after 30 min. In most memory experiments, the saving score is the difference between the number of repetitions required to learn the material and the number required for re-learning, expressed as a percentage of the former. In this case, the formula was

$$\frac{N_1 - N_2}{N_1} \times 10 + 5,$$

where N_1 = number of repetitions for learning, and N_2 = number of repetitions for re-learning (summed over all the items). The constant was added in order to make all the scores positive as, in a few cases, N_2 was larger than N_1.

4) *Immediate memory*

The main object in using an immediate memory score was to link the present research with previous work, most of which has made use of this type of score. It was equal to the number of items recalled or recognized correctly on the first attempt in the learning phase.

The main characteristics of all the memory tests are summarized in Table 1.

IV. *Results*[1]

The first stage in the analysis was the computation of product-moment correlation coefficients between all the test scores.[2] For the purpose of this investigation 'g' was defined as the first general factor obtained by an analysis of the correlations between all the Wechsler subtests, using Burt's Simple Summation method. In this analysis one

[1] I am indebted to Mr A. Lubin who gave invaluable assistance in the analysis of the results.

[2] In the case of the Wechsler, the subtest raw scores were used for this purpose.

general factor and two bi-polars were extracted, using estimated communalities and successive approximations.

Three criteria were considered in determining the number of factors to be extracted. The first was the comparison of the product of the two highest factor loadings with the standard error of a zero correlation (4). The second was the graphing of the sums of residuals (excluding signs) against the number of factors extracted (7). The third was the comparison of residuals with the correlations which might be expected to arise by chance, assuming a true correlation of zero. Using the first two criteria, the third factor was found to be significant whilst the fourth was not. Using the third criterion, the third factor was not significant. The third factor was retained on the grounds that the latter criterion has been shown to result in the extraction of too few factors (9).

Table 1. *Summary of memory tests*

Test	Score	No. of pairs	Response items	Method
A	11 Learning 12 Retained I 13 Saving 31 Immediate M	8	Words	Recall
B	14 Learning 15 Retained I 16 Saving 33 Immediate M	8	Nonsense syllables	Multiple choice
C	17 Learning 18 Retained I 19 Saving 34 Immediate M	8	Objects	Multiple choice
D[1]	32 Immediate M	3	Nonsense figures	Recall
E[1]	20 Learning 21 Retained I 35 Immediate M	8	Nonsense syllables	Recall
F	22 Learning 23 Retained I 24 Saving 36 Immediate M	8	Objects	Recall
G[1]	25 Learning 26 Retained I 37 Immediate M	10	Words	Multiple choice
H	27 Learning 28 Retained I 29 Saving 38 Immediate M	8	Nonsense figures	Multiple choice

[1] Certain scores omitted because some of the subjects were unable to complete the tests.

To determine whether the memory test intercorrelations provided evidence of a memory factor in addition to '*g*', residuals were computed, with the effect of '*g*' eliminated. As described above, '*g*' was defined as the first general factor in the analysis of the correlations between Wechsler subtests. The loadings of the memory tests on this factor were calculated from the correlations between memory tests and the Wechsler subtests, using the method described by Thurstone (8) in the Grouping Method of Factoring. The residuals were then derived in the usual way from the memory test intercorrelations. These residuals are shown in Table 2.

To show that '*g*' was not sufficient to account for the correlations between memory tests, it was necessary to establish that the residuals could not be attributed to chance.[1] The average figure in Table 2, retaining the signs, is 0·194 and, to get an objective

[1] Those residuals resulting from different scores from any one test are excluded from the following calculations, as they would tend to raise the level of significance spuriously.

Table 2. *First-factor residuals* (*memory tests*)

		A				B				C				E			F				G			H				D
		11	12	13	31	14	15	16	33	17	18	19	34	20	21	35	22	23	24	36	25	26	37	27	28	29	38	32
A	11	—																										
	12	504	—																									
	13	248	709	—																								
	31	598	449	234	—																							
B	14	086	109	187	−015	—																						
	15	135	136	102	−018	322	—																					
	16	−025	061	060	−133	032	707	—																				
	33	062	106	157	−029	541	411	150	—																			
C	17	190	164	099	039	271	075	087	175	—																		
	18	183	119	089	−016	093	073	192	084	451	—																	
	19	068	043	090	005	020	093	247	007	081	697	—																
	34	108	085	018	052	201	061	115	112	798	433	153	—															
E	20	050	107	150	009	275	212	178	140	115	087	067	100	—														
	21	228	290	271	188	214	174	068	124	003	023	051	−001	320	—													
	35	116	115	054	248	−010	−062	−035	−112	080	005	−038	045	355	198	—												
F	22	039	022	102	131	031	051	089	−097	084	199	163	081	211	011	134	—											
	23	259	235	226	230	−103	086	155	−129	−015	238	242	074	023	028	084	156	—										
	24	262	243	210	280	−088	056	029	−041	−007	102	087	093	−046	078	145	−177	715	—									
	36	161	103	108	275	−003	077	158	−099	−045	087	139	025	126	031	126	717	207	−102	—								
G	25	269	195	177	166	163	289	272	116	128	113	104	093	287	238	151	259	088	028	246	—							
	26	368	358	304	404	114	162	150	074	056	092	173	067	265	350	307	190	146	121	349	389	—						
	37	255	265	220	323	103	228	198	096	074	043	101	082	180	187	124	245	110	074	354	562	507	—					
H	27	−133	−225	−128	−222	031	−058	001	−023	014	124	148	036	010	040	−126	163	−047	−071	025	015	−094	−027	—				
	28	033	−027	035	024	176	021	−060	167	056	188	215	031	097	109	−028	064	−015	−043	075	−113	136	002	108	—			
	29	064	077	123	131	210	028	−079	130	108	208	256	057	042	109	−034	052	058	−012	108	−106	154	−028	−167	653	—		
	38	124	087	042	025	056	062	093	000	078	112	081	127	089	168	−040	043	−037	−079	−006	131	093	096	485	088	−157	—	
D	32	205	222	177	246	081	−024	−058	−027	015	084	043	041	021	178	094	221	090	007	248	065	349	168	067	221	177	098	—

criterion of significance, the observed distribution may be compared with the distribution of correlations which might be expected to arise by chance in samples from a population where the true correlation is zero. Strictly speaking, the residuals should be converted to partial correlations to make this comparison but, as the conversion can only increase the level of significance, it will be unnecessary if the criterion shows the unconverted residuals to be significant. Table 3 shows the coefficients which would be expected to appear or be surpassed in 10, 5, 2 and 1% of samples from the population specified by the null hypothesis. This is for samples where $N = 82$, and is taken from Fisher's table of values for different levels of significance (3). Also shown in this table are the percentages of residuals which actually reach or surpass these values in Table 2.

An examination of Table 3 shows that fewer high positive coefficients would be expected to arise by chance, if the true correlation was zero, and more high negative ones. The difficulty which underlies this criterion of significance is that it is only strictly applicable when the correlations have been obtained from independent samples of the

Table 3. *Chance and observed distributions of residuals above* 0·18

r	0·18+	0·22+	0·26+	0·28+
Expected by chance (%)	10	5	2	1
Observed (positive) (%)	21·0	14·3	7·3	4·4
Observed (negative) (%)	0·6	0·6	0·0	0·0

same population. When, as in this investigation, the correlations are obtained by giving different tests to the same group, the sampling errors of the correlation coefficients are not independent, the shape of the distribution is unknown and the use of the test for most purposes cannot be justified. In this case, however, the criterion was selected for a definite reason. It has been found that this method, used to determine the number of factors to be extracted from a correlation matrix, tends to err in the direction of too few factors rather than too many. In fact, it seems invariably to result in the extraction of fewer factors than any of the other well-known methods (9). The fact that this criterion shows that the residuals cannot be accounted for by chance, makes it safe to assume that the first condition necessary for the existence of a factor common to the memory tests, other than '*g*', is established. A further indication will be found in the factor loadings.

The next stage in the analysis was to calculate the loadings of the memory tests on this factor, from the residuals shown in Table 2. To overcome the problem of the spurious correlations between different scores from the same tests, Burt's Group-factor method of analysis was employed, assuming the existence of one general factor (*m*) and seven group factors, each common to all the scores from any one test (2). Using this method, the diagonal submatrices which may be seen from Table 2 to contain all the spurious residuals, do not enter into the calculation of the first factor loadings. The loadings of the memory tests on '*g*' and '*m*' are shown in Table 4. Applying Guilford's criterion for the significance of factors, the product of the two highest '*m*' loadings (0·693 and 0·513) is 0·356 which is well in excess of the S.E. of a zero correlation and confirms the conclusion based on the size of the residuals. The fact that only one learning score does not have a positive loading on this factor and that only two have loadings of less than +0·15, indicates that this factor is common to a large range of memory tests.

An examination of the residuals after the extraction of '*m*' showed that there was little likelihood of further significant group factors, apart from those common to different

scores from the same test. Excluding the figures in the diagonal submatrices, only eleven entries lay outside the range $\pm 0{\cdot}18$ and the highest was 0·197.

V. Discussion

The results, as presented so far, lead to the following conclusions:

(i) That '*g*' does not account adequately for all the correlations between memory tests.

(ii) That '*m*', having positive loadings for all the memory scores, except one Learning score, is general as far as these memory tests are concerned. It may be interpreted, tentatively, as a 'memory' factor.

Table 4. *Memory test loadings on 'g' and 'm'*

Test		Score	'*g*'	'*m*'
A	11	Learning	0·382	0·455
	12	Retained I	0·373	0·423
	13	Saving	0·333	0·420
	31	Immediate M	0·283	0·343
B	14	Learning	0·518	0·291
	15	Retained I	0·300	0·259
	16	Saving	0·130	0·233
	33	Immediate M	0·421	0·131
C	17	Learning	0·267	0·253
	18	Retained I	0·323	0·345
	19	Saving	0·296	0·329
	34	Immediate M	0·307	0·233
E	20	Learning	0·499	0·373
	21	Retained I	0·405	0·422
	35	Immediate M	0·525	0·179
F	22	Learning	0·363	0·341
	23	Retained I	0·415	0·278
	24	Saving	0·147	0·196
	36	Immediate M	0·302	0·366
G	25	Learning	0·493	0·499
	26	Retained I	0·372	0·693
	37	Immediate M	0·472	0·513
H	27	Learning	0·485	−0·060
	28	Retained I	0·466	0·172
	29	Saving	0·200	0·231
	38	Immediate M	0·290	0·182
D	32	Immediate M	0·208	0·371

When the precise nature of the memory factor is considered, there are two possible interpretations. All the test scores, including retained items and, probably, saving, involve the ability to memorize the material (owing to the fallibility of the learning criterion). It is conceivable, therefore, that the common factor may be due entirely to techniques of memorization, applicable over this range of material. This was essentially the conclusion reached by Anastasi (1) on the nature of the factor that she found common to a group of tests involving immediate memory for rote material. The finding which led her to this conclusion was that the same factor did not appear in a group of tests covering a wider range of memory abilities, including delayed recall. If this interpretation should be correct, then clearly '*m*' would be of very limited importance in determining intellectual abilities.

An alternative possibility, however, is that '*m*' is a genuine 'retentivity' factor, affecting the retention phase of the memory process, rather than the phase of initial impression. It seemed very desirable that an attempt should be made to decide which of these alternative explanations was likely to be the correct one, and the design of the investigation led to a method of making such an attempt.

The Retained items and Saving scores were designed to exclude, as far as possible, any variation due to initial learning level. Thus, any factor affecting only the learning phase of the memory process would be expected to have a minimum influence on these two scores. The Learning and Immediate memory scores, on the other hand, would be expected to depend to a considerable extent on factors affecting only initial learning. If '*m*' represented a genuine 'retentivity' factor, therefore, it would be expected to have more influence upon Retained items and Saving, whereas, if it should be 'memorization', it would be expected to have more influence upon Learning and Immediate memory. As nothing was known about the error and specific variance of each test, it was not sufficient to compare the loadings on '*m*' directly. What we could do, however, was to compare the relative contributions to the variance of each test of '*g*' and '*m*'. On the 'memorization' hypothesis '*m*' would be expected to contribute more to the common factor variance of the Learning and Immediate memory scores than to that of the Retained items and Saving scores. On the 'retentivity' hypothesis, however, the reverse would be the case.

Table 5

	Percentage of variance attributable to	
	'*g*'	'*m*'
Learning scores	19·2	12·6
Retained items	14·6	16·1
Saving scores	5·5	8·6
Immediate memory	13·4	9·9

Table 5 shows the figures necessary to make this comparison, namely, the average percentages of variance accounted for by the two factors, in each of the four sets of scores. ($100\,[\Sigma k^2]/N$, where Σk^2 = sums of squares of factor loadings, and N = number of variables.)

It is at once evident that the results support the hypothesis that it is the retention phase of the memory process which is influenced by '*m*'. Both Retained items and Saving scores depend more upon '*m*' than upon '*g*', whilst both Learning and Immediate memory scores depend more upon '*g*'. It seems reasonable, therefore, to refer to '*m*' as 'retentivity'. An important defect in this argument is that there is no test of significance of the differences between the factor variances. For this reason, it is only possible to say that the results 'support' the hypothesis, and it is only with repetition of the findings that the degree of confidence with which the assertion is made can be increased.

Earlier, it was suggested that a possible explanation of the fact that evidence confirming the existence of a unitary trait of memory was lacking, was that tests of Immediate memory had been used in most investigations. The comparative factor variances reported above support this suggestion. It is in tests of this type that 'retentivity' would be least likely to appear as a common factor.

Because such a factor has been defined and found common to a certain group of memory tests, it does not follow automatically that it includes all tests where memory is involved. For our knowledge of the 'retentivity' factor to be in any way complete, we should be able to state its extent, and only if it covers all memory tests is it possible to refer to it as a general memory factor.

Is it possible, therefore, to establish the extent of the memory factor by the results of this investigation? The tests cover a wide range, including verbal, pictorial, nonsense

and meaningful material. They include both recall and multiple choice methods of testing retention, and assessments are made of several phases of the memory process from immediate memory to memory after 30 min. It must be admitted, however, that the tests certainly do not cover the whole range of abilities which might involve retentivity. Probably the greatest deficiency is that none of the material is meaningful in the sense of having an intrinsic organization, whereby the different items are related. It is likely that the ability to retain a passage of prose, for instance, or the solution to a complex problem, involves quite different functions from those underlying retention for rote material, and it has often been stressed that meaningful retention plays a far greater role in practical living than rote memory. It might be argued, therefore, that a factor which only influenced meaningless material would be of merely academic interest.

To counter this argument, it might be suggested: (*a*) that rote learning does cover a considerable proportion of the learning that takes place in schools and, therefore, must be of practical importance; and (*b*) that the distinction between meaningful and meaningless material is merely one of degree and that the two forms differ only in the amount of organization present in the material. In this case, as some degree of organization enters into all learning processes, there is no reason why the most highly organized should involve essentially different functions from the least highly organized. Clearly, the only way to settle the question is by further investigation.

All that can be concluded, at present, is that in a sample of male adults, a factor has been defined which appears to be most reasonably interpreted as retentivity for unrelated material. The extent to which it is involved in other intellectual abilities, including school achievement and memory for related material, is a matter for future research to decide. To answer the question with respect to school achievement, it would be necessary to include some of these memory tests in a battery with measures of academic success and the same could easily be done with criteria of job proficiency. Similarly, a research could now be planned to include with some of the present memory tests, others involving memory for related material.

VI. Summary

Eight memory tests were devised, each one involving initial learning up to a constant criterion and subsequent re-learning after 30 min. Each test was scored in four different ways to assess: (*a*) learning ability, (*b*) immediate memory, (*c*) retentivity measured by retained items, and (*d*) retentivity measured by saving. The time interval was occupied by subtests from the Wechsler Bellevue Intelligence Scale which was used to define '*g*' for the purpose of the investigation. Product-moment correlations were computed, based on the scores of eighty subjects.

A factor-analysis of all the correlations indicated that a factor, other than '*g*', was required to account for the correlations between memory tests. The results showed that this factor could reasonably be interpreted as one influencing the retention phase of the memory process.

References

(1) Anastasi, A. (1932). Further studies on the memory factor. *Arch. Psychol.* xxii no. 142.
(2) Burt, Sir C. (1940). *The Factors of the Mind.* London: University of London Press.
(3) Fisher, R. A. (1946). *Statistical Methods for Research Workers,* 10th ed. Edinburgh: Oliver and Boyd.
(4) Guilford, J. P. (1947). Printed classification tests. *Army Air Forces Aviat. Psychol. Progr. Res. Rep.* no. 5. Washington, D.C.: U.S. Government Printing Office.
(5) Ingham, J. G. (1949). An investigation into the relationship between memory and intelligence. Unpublished thesis. University of London.
(6) McGeoch, J. A. (1942). *The Psychology of Human Learning.* New York: Longmans, Green and Co.
(7) Mosier, C. I. (1939). Influence of chance error on simple structure: an empirical investigation of the effect of chance error and estimated communalities on simple structure in factorial analysis. *Psychometrika,* iv, 33–44.
(8) Thurstone, L. L. (1947). *Multiple-Factor Analysis.* Chicago: University Press.
(9) Vernon, P. E. (1949). How many factors? Unpublished Memorandum, The Admiralty, London.
(10) Wechsler, D. (1946). *The Wechsler-Bellevue Intelligence Scale: Form II.* Manual for administering and scoring the test. New York; the Psychological Corporation.
(11) Woodworth, R. S. (1914). A contribution to the question of 'quick learning, quick forgetting'. *Psychol. Bull.* xi, 58–59 (Abstr).

(*Manuscript received* 27 *February* 1951)

TEMPERAMENTAL DIFFERENCES IN MAZE PERFORMANCE

PART II. THE EFFECT OF DISTRACTION AND OF ELECTROCONVULSIVE THERAPY ON PSYCHOMOTOR RETARDATION

By G. A. FOULDS, *Runwell Hospital*

I. *Introduction* (p. 33). II. *Results of immediate re-test on Mazes:* (i) *with distraction* (p. 35); (ii) *without distraction* (p. 36). III. *Maze results of reactive depressives and melancholics* (p. 38). IV. *Maze results before and after E.C.T.* (p. 39). V. *General summary* (p. 40). *References* (p. 41).

I. Introduction

The Porteus Maze test, with certain modifications in its scoring and administration, was found to differentiate between subjects suffering from various forms of psychoneurotic illness. One of the principal findings was that reactive depressives, anxiety states, and obsessionals tended to work more slowly than did hysterics and psychopaths.

In the somewhat barren literature on the subject of retardation MacCurdy's work(7) still stands out, despite some apparent inconsistencies, and still represents perhaps the most widely held point of view. 'Retardation', says MacCurdy, 'is a direct expression of deficient psychic energy. Painful thoughts are slow. Often in our daily life the thought of something unpleasant suddenly "blocks" our mental processes and we are temporarily incompetent. Such a sudden inhibition is a miniature depression—when *any* personal topic produces such an interference with thought, the reaction has developed to a definitely psychotic degree.' He also writes: 'This phenomenon of preoccupation with conative incompetence suggests that mental activity is not cut off at its source, that it is not non-existent, but rather that it is constantly striving for expression against some inhibitive force which persistently represses it.' Again, 'retardation seems to be a blocking in response to milder and milder occasions for mental pain as the depression increases in gravity...the preoccupation of the patient's attention with his disability suggests that repression is in operation.'

Retardation is introduced first in terms of deficient psychic energy—the decelerating machine analogy—but this view is not maintained. 'The thought of something unpleasant suddenly blocks our mental processes' heralds the confusion. It is only previous, or constructive thoughts presumably that are blocked. Some mental activity is still going on. There is no void, since we are thinking of something unpleasant; in fact we are preoccupied with our 'conative incompetence', which suggests to MacCurdy that mental activity 'is not cut off at its source'. MacCurdy seems sometimes to have forgotten that these preoccupations are themselves mental activities presumably using up 'psychic energy'. It is not that painful thoughts are necessarily slow, but that other more constructive thoughts are slowed up by the intrusion of painful thoughts. An agitated melancholic may be extremely fluent in recounting the supposed causes of her anguish. It is when she is asked to do something else that she may become slow. Again, do not the preoccupations suggest that displacement rather than repression is in operation?—though, of course, displacement presumes repression. Attention and affect have perhaps been displaced from the original guilt-laden situation to the avowed inability to con-

centrate, to the failure of 'memory', to the unspecified wickedness, to the commission of the undefined but unforgivable sin.

Wiener suggests that the persistent worry of the anxiety neurotic may use up so much of his neuron pool that he is relatively incapacitated from carrying out other processes of thought. The vicious circle that the cyberneticists call positive feed-back may even incapacitate the individual almost completely, more and more of the neurons being incorporated in the system(13). The more extreme incapacity would presumably be seen in cases of agitated melancholia.

Despite his caustic comments on physical treatments in psychiatry, Wiener—perhaps because of his adherence to a mechanistic viewpoint—does not appear to envisage any means of intervening in a circulating system other than by violent physical assault. It will be shown that such a means does exist. Explanations of retardation in terms of oscillations of attention, breaking up of behaviour patterns, availability of neurons, etc., are not necessarily incompatible, but may be partial and biased.

It is widely recognized that fear is experienced before and after, rather than during, a dangerous activity which engages the entire attention. A rugby footballer may sustain a cut without either becoming aware of it or feeling any pain until the end of the game. Less strenuous pursuits may have a similar effect; the pain-threshold may, in fact, be raised by as much as 45% if the subject be asked to memorize and repeat numbers(14). The Roman Catholic tells his rosary to help exclude distracting thoughts. Distracting attention from the visual to the auditory field may affect the electrical activity of the brain; for example, a subject wearing plus 10D lense spectacles found the visual field 'uninteresting', but the alpha rhythm showed some blocking. A ticking watch close to the ear brought about a reappearance of the alpha rhythm(1).

Other evidence on the importance of the value of the stimulus to the subject is found in Weber's report(12) that some of the most considerable losses in output occurred when pleasant distracting stimuli were used, such as beautiful music or humorous anecdotes. The significant feature is, however, the compelling nature of the interest rather than its feeling-tone. The depressed patient's brooding over his sinful past will be much more distracting to him than a Mozart symphony.

When confronted with the Maze test situation not one patient suffering from an anxiety state has yet been either unwilling or unable to do the test in some fashion. The new behaviour pattern, the test situation, might, therefore, be said to break up the more habitual pattern, the recurrent anxiety. Somehow the neurons are made available. The subject can attend, at least intermittently, to the behaviour pattern which he has tacitly accepted. In approximately 10% of cases of agitated melancholia the subject is quite unable to do the test. The habitual behaviour pattern cannot be disrupted for a sufficiently long time. He, or more probably she, cannot attend to the presented situation. Possibly no neurons are available.

If the Maze test fails to absorb sufficient of the dysthymic's attention to exclude affective competition, engagement in a subsidiary task may temporarily bring about such an exclusion. In the experiments to be reported the intention was to begin with minimal distraction and, if this proved ineffective, to increase the complexity of the distracting stimuli until some observable change in performance was produced.

Previous work on distraction has been mainly concerned with the effect of external stimuli such as loud noises, flashing lights, etc., on speed and accuracy of performance

at a given task (5, 11), sometimes coupled with an assessment of energy expenditure (6, 9). Where the subject has engaged actively in both tasks, the original activity has been of a quasi-automatic type (3). The majority of these investigations have been carried out with normal subjects; with mentally disturbed patients there would be no certainty that such external stimuli as lights and noises had been perceived at all. It was, therefore, preferable that in the present investigation the subject should engage in some overt activity.

Clearly the effect produced will vary with the relative simplicity or complexity of the two tasks. Since maze tracing is a relatively complex task, it was considered that the present need would best be met by a simple subsidiary activity and, since intermittent distracting stimuli produce more effect than continuous stimuli, the latter kind was preferred.

Three conditions, therefore, seemed desirable for minimal but certain distraction: (i) that the subject should engage in some overt activity; (ii) that the distracting stimulus should be simple; (iii) that it should be regular. Accordingly, after performance of the test in the manner described in Part I of this paper (4), the subjects were asked to perform the test again under the same conditions with the one exception that they were to repeat numbers after the tester. The tester counted 1, 2, 3, etc., at intervals of approximately 2 sec.

It was predicted that: (i) without distraction in either test, immediate re-test would result in an increase in speed of performance on the second test in both the dysthymic and non-dysthymic groups as a consequence of increased familiarity or practice; (ii) with distraction on the second test, immediate re-test would result in a significantly greater speed of performance with the dysthymic group on the supposition that the counting would temporarily obscure the anxiety or depression, or break up this behaviour pattern and that practice effect would be operating in the same direction; (iii) with distraction on the second test, immediate re-test would result in no significant change in speed of performance with the non-dysthymic group on the supposition that the counting would cause a partial deflexion of attention which would operate in a direction opposite to that of the practice effect.

II. Results of immediate re-test on Mazes

(i) *With distraction*

Seventy-five psychoneurotics were tested, fifteen in each of the subgroups. The ordinary form was given first and the distraction form immediately afterwards in the same testing session. Table 1 shows the scores on the distraction form of the test expressed as a percentage of the scores on the ordinary form. The features of the performance here scored separately are Starting Time, Tracing Time, Weighted Wrong Direction Score, Lifted Pencils, Crossed Lines, Wavy Lines (cf. (4)).

With two minor exceptions, psychopaths and hysterics are again clearly differentiated from the dysthymics. The former group was, however, younger. The fifteen oldest dysthymics were therefore discarded, and Table 2 gives the results of the remaining thirty dysthymics and thirty non-dysthymics. The dysthymic group had a mean age of $32{\cdot}37 \pm 6{\cdot}91$ and a Progressive Matrices (1938 untimed) score of $41{\cdot}00 \pm 9{\cdot}00$; the corresponding figures for the non-dysthymic group were $31{\cdot}13 \pm 9{\cdot}62$ and $41{\cdot}23 \pm 11{\cdot}24$.

Starting times are given in seconds, tracing times in quarter minutes in Tables 1, 2, 3 and 4.

The critical ratio between the means of the dysthymics and non-dysthymics on the ordinary form indicated differences significant at least at the 5% level of confidence on three of the scoring categories; whereas, on the distraction form, there were no such significant differences. The relevant critical ratios were: starting time: ordinary form 1·99; distraction form 0·36; tracing time: 3·07 and 0·66; crossed lines 2·04 and 1·42.

In order to determine whether or not the subjects who took longest, irrespective of

Table 1. *Scores on distraction form as percentage of score on ordinary form for seventy-five psychoneurotics*

	St.T.	T.T.	W.W.D.	L.P.	C.L.	W.L.
Hysterics	75	89	118	66	136	111
Psychopaths	80	91	127	79	124	95
Anxiety states	75	69	111	45	169	91
Depressives	47	64	123	56	164	97
Obsessionals	52	71	110	42	171	90

Table 2. *The ordinary and distraction forms of the Mazes with thirty dysthymic and thirty non-dysthymic subjects*

	Dysthymics		Non-dysthymics	
	Mean	S.D.	Mean	S.D.
Ordinary form:				
St.T.	56·10	57·72	33·87	20·19
T.T.	35·87	18·27	24·57	8·52
W.W.D.	18·90	11·81	19·73	12·60
L.P.	7·40	6·81	6·37	6·56
C.L.	7·80	9·52	13·23	11·00
W.L.	2·30	2·18	2·47	1·89
Distraction form:				
St.T.	27·97	19·46	26·37	15·21
T.T.	23·50	8·31	22·07	8·39
W.W.D.	22·87	13·36	24·40	12·31
L.P.	3·87	5·12	4·60	5·19
C.L.	12·47	13·40	16·90	10·51
W.L.	2·20	1·89	2·53	1·94

type of illness, show the greatest increase in speed under distracting conditions, the ten slowest dysthymics and the ten quickest non-dysthymics were discarded. The tracing times of the remaining subjects were then compared. The means on the ordinary form were: dysthymics $29{\cdot}00 \pm 8{\cdot}07$; non-dysthymics $28{\cdot}95 \pm 7{\cdot}04$; on the distraction form the respective means were $21{\cdot}40 \pm 6{\cdot}22$ and $25{\cdot}90 \pm 7{\cdot}56$. Whereas the difference between the means for the two groups on the ordinary form of the test is not at all significant, the critical ratio of the difference between the means on the distraction form is 2·05.

While, therefore, there may possibly be a slight tendency for the gain in speed to be greatest with those who were initially very slow, the presence or absence of emotional disturbance is clearly of much greater importance. Presumably, therefore, many dysthymics would, but for their illness, have performed unusually quickly.

Predictions (ii) and (iii) in § I are, therefore, supported by the results of this part of the experiment.

(ii) *Without distraction*

Another thirty dysthymics and thirty non-dysthymics were given the non-distraction form of the test twice in the one session. The dysthymic group had a mean age of

37·20 ± 12·57 and Progressive Matrices score of 37·93 ± 13·09; the corresponding figures for the non-dysthymic group were 34·93 ± 12·22 and 38·77 ± 11·04.

Table 3 shows the scores on the non-distraction form of the test and on re-test with the same form.

Table 3. *Scores for two performances of the non-distraction Mazes with thirty dysthymic and thirty non-dysthymic subjects*

	Dysthymics		Non-dysthymics	
	Mean	S.D.	Mean	S.D.
Ordinary form:				
St.T.	55·07	47·25	46·77	46·17
T.T.	37·33	16·27	29·67	13·54
W.W.D.	21·20	15·09	18·40	8·76
L.P.	13·20	12·83	7·07	6·61
C.L.	9·60	10·14	11·47	11·20
W.L.	1·97	2·20	2·20	2·41
Re-test with ordinary form:				
St.T.	42·63	35·26	34·77	33·46
T.T.	31·93	16·74	22·90	7·43
W.W.D.	18·27	12·38	16·50	7·73
L.P.	8·37	8·86	5·80	6·66
C.L.	11·80	10·67	13·23	9·77
W.L.	1·73	1·96	1·73	2·06

Table 4 shows the St.T. and T.T. scores (recorded in Table 3) on the second test expressed as a percentage of the scores on the first test and the St.T. and T.T. scores (recorded in Table 2) on the second test as a percentage of the scores on the first test. O = ordinary form; R = re-test on ordinary form and D = distraction form.

Table 4. $\frac{\text{R (St.T. and T.T. scores)}}{\text{O (St.T. and T.T. scores)}} \times \frac{100}{1}$ *and* $\frac{\text{D (St.T. and T.T. scores)}}{\text{O (St.T. and T.T. scores)}} \times \frac{100}{1}$ *for two groups of thirty dysthymics and thirty non-dysthymics*

	Dysthymics				Non-dysthymics			
	$\frac{R}{O} \times \frac{100}{1}$		$\frac{D}{O} \times \frac{100}{1}$		$\frac{R}{O} \times \frac{100}{1}$		$\frac{D}{O} \times \frac{100}{1}$	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
St.T.	89·50	37·19	66·33	29·32	86·97	31·32	91·03	74·69
T.T.	84·03	18·55	70·30	18·75	82·83	27·27	91·43	22·77

From this table it can be seen that the effect of immediate re-test with the non-distraction form was to increase the speed of performance in both groups to an equal extent. The following effects of distraction on Maze performance may, therefore, be noted: (i) an increase in speed of work with dysthymics, the mean of the difference between the O:R and O:D ratios being significantly different from zero (for St.T.'s, $t = 2{\cdot}27$; for T.T.'s, $t = 2{\cdot}61$); (ii) no increase in speed of work with non-dysthymics. There is, in fact, a slight, though not significant, decrease.

Other possible effects of distraction which require further confirmation are: an increase with both groups in the number of errors of direction and in the number of crossed and wavy lines; a decrease with both groups in the number of lifted pencils.

Morgan (9) found that a typist submitted to a barrage of noises could actually increase speed of work without loss of accuracy. Physiological recordings indicated that more energy was expended on the performance. The original activity was, however, a quasi-habitual one, so that one would expect considerable reserves of energy to be available.

This is much less true in the case of the generally unfamiliar and complex task of maze tracing, so that the considerable changes in the performance of the dysthymics are not likely to be attributable merely to an increase in effort. This seems particularly so in that the most severely depressed or obsessional patients drive themselves to the point of desperate striving even at the first performance. There is probably little reserve of effort left in such cases.

If the present results be more adequately accounted for by the assumption that retardation may be largely due to partial deflexion of attention, the technique described may contribute towards the formidable task of measuring the depth and intensity of the affective disturbance and towards further elucidation of the concept of retardation. Psychiatric discussion of the problem of retardation has remained, for the most part, at the purely descriptive level. Even if, in spite of the findings of Cameron (2), McFarland & Goldstein (8) and others, psychomotor retardation be treated as an effect of metabolic or other physical changes, then these changes are of such a nature that the effect can be temporarily offset by purely psychological means of a very simple type.

Further evidence could be provided for or against the hypothesis put forward in this paper by testing subjects suffering from a depressive illness both before and after a course of, for example, electroconvulsive therapy.

Before E.C.T. the starting and tracing times under distraction should be considerably shorter than for the non-distraction form. If E.C.T. blots out or breaks up the depressive pattern, the discrepancy between the distraction and non-distraction forms should be greatly reduced. The suggestion is, therefore, that the effect of counting 1, 2, 3 and of a course of electroconvulsive therapy have at least one important feature in common—the weakening or breaking up of certain behaviour patterns promoted by affective disturbances. Since E.C.T. has a somewhat more enduring effect, it may allow older behaviour patterns to be re-established. Prolonged distraction techniques such as Occupational Therapy (under which should be included the latest fissions, Social and Art therapy) may produce a similar result.

Since the previous depressive groups were all psychoneurotics and since the majority receiving E.C.T. are psychotic, two such groups were first tested to determine whether or not there was any essential difference in their performances.

III. Maze results of reactive depressives and melancholics

Each group consisted of seventeen subjects matched for age, sex and Progressive Matrices score. The reactive depressives included five men, the mean age was 49·3 years (S.D. 9·8), the mean Matrices score 30·9 (S.D. 11·9); the melancholics included three men, the mean age was 49·8 years (S.D. 9·7), the mean Matrices score 31·3 (S.D. 9·2). The two groups were given the Porteus Maze test without and then with distraction. The results are shown in Table 5.

Only one difference between the two groups was significant at the 5% level of confidence. On the ordinary form of the test the melancholics were slower on tracing time. They were slower on the other three speed measurements, but the spread of scores was very great. None of the six remaining scoring categories produced a significant difference. The effect of distraction was similar for both groups in all categories. All the characteristics of the performance of reactive depressives are, therefore, applicable to melancholics, usually in a more marked degree.

IV. Maze results before and after E.C.T.

Thirty subjects were tested and the group was made up of twelve melancholics, two manic-depressives (in the depressive phase), eleven anxiety-depressives and five cases in which the differential diagnosis between psychotic and psychoneurotic depression was uncertain. There were twenty-three women and seven men. Both before and after E.C.T. the ordinary form of the test preceded the distraction form, each being given at the same session. The results are shown in Table 6.

Table 5. *Maze results without and with distraction of seventeen reactive depressives and seventeen melancholics*

	Reactive depressives		Melancholics	
	Mean	S.D.	Mean	S.D.
Ordinary form:				
St.T.	54·71	43·04	120·53	164·02
T.T.	47·65	18·87	70·53	40·61
W.W.D.	24·35	12·25	22·82	18·63
L.P.	15·65	12·84	27·59	34·23
C.L.	8·76	7·40	12·71	12·31
Distraction form:				
St.T.	28·88	11·50	45·06	37·21
T.T.	32·24	24·52	44·12	21·72
W.W.D.	25·82	11·94	26·47	18·54
L.P.	9·47	12·18	16·76	20·62
C.L.	11·59	11·39	14·88	10·91

Table 6. *Results of the ordinary and distraction forms of the Mazes before and after E.C.T.*

	Ordinary form before E.C.T.		Distraction form before E.C.T.	
	Mean	S.D.	Mean	S.D.
St.T.	56·00	41·49	33·04	21·38
T.T.	53·27	33·00	37·06	21·44
W.W.D.	26·87	15·14	27·27	14·39
L.P.	23·80	27·70	14·13	16·13
C.L.	12·60	8·97	14·60	10·73
	Ordinary form after E.C.T.		Distraction form after E.C.T.	
	Mean	S.D.	Mean	S.D.
St.T.	43·83	36·03	35·40	33·13
T.T.	38·37	18·88	33·17	16·30
W.W.D.	19·77	11·44	22·17	11·47
L.P.	6·83	6·87	5·13	6·97
C.L.	9·10	8·77	10·23	7·81

Other tests were also administered as a routine matter. The mean pre- and post-E.C.T. scores on the Mill Hill Vocabulary Scale are virtually unchanged; allowing for practice effect, this is equally true of the Progressive Matrices scores. The important difference lies in the time taken to complete this test. Before E.C.T. twenty subjects and after E.C.T. only four subjects took more than an hour to do Progressive Matrices. On the average, subjects reduced the time from approximately 100 to approximately 50 min.

On the Maze test distraction has presumably a slightly unfavourable effect on the W.W.D. score counteracting the small practice effect noted in Table 3. The improvement after E.C.T. is not significantly greater than might be expected merely from practice

effect (i.e. the mean of the difference between the two W.W.D. scores of dysthymics in Table 3 is not significantly less than the mean of the difference between the W.W.D. scores on the ordinary form pre- and post-E.C.T. The C.R. of 1·73 is, however, suggestive).

There is a steady reduction in the number of lifted pencils with each performance. Both distraction and re-test have been shown to improve this score. To these two facts may perhaps be added a lessening of the depression after E.C.T., which would fit in with earlier results on depressives and normals and with the continued and statistically significant drop in L.P. score between tests 2 and 3 ($t=2·89$).

The Crossed Lines score is slightly increased with distraction, which is in keeping with previous results. There is, however, a substantial drop in C.L. score after E.C.T., since re-test without E.C.T. has been shown (Table 4) to result in an increased C.L. score.

The effect of distraction on the starting and tracing times before E.C.T. is as expected from the results shown in Tables 2 and 5. Times on the ordinary form after E.C.T. are similar to those on the distraction form before E.C.T.—the mean St.T. would have been considerably closer but for one exceptionally slow subject. Distraction certainly does not have the same speeding-up effect after E.C.T. as it has before; the mean of the difference between the O:D ratios pre- and post-E.C.T. being significant for both St.T. and T.T. (for St.T., $t=2·50$; for T.T., $t=4·44$). This finding supports the hypothesis which this experiment was designed to test.

It may be concluded, therefore, that the effect of distraction on speed of performance in the Maze test is a useful method of assessing the intensity of an affective disturbance.

Sargant(10) writes that, after E.C.T., 'thought rumination is dispersed; recently formed attitudes lose their hold on the patient, and older established patterns of thinking are restored with the abolition of those often diametrically opposed to them which supervened during the illness'. The experiments now reported indicate that distraction may have a similar though, in the present circumstances, temporary effect. It may achieve this effect by drawing the attention away from the affective disturbance; whilst E.C.T. in some way reduces the intensity of the unpleasant affect and thus enables the activity with which it is competing to dominate consciousness more frequently than had previously been possible.

V. General summary

1. Certain modifications were made in the administration and scoring of the Porteus Mazes with a view to emphasizing the temperamental factors involved in its performance. Particular attention was given to speed of performance.

2. Differences in performance due to sex, age and psychoneurosis were studied, and a comparison made between the results obtained from subjects suffering from different forms of psychoneurotic illness.

3. The test was found to differentiate more clearly between psychoneurotic groups than between psychoneurotics and normals.

4. The characteristic performances of psychopaths, hysterics, anxiety states, obsessionals and reactive depressives were described. Speed of performance decreased in that order.

5. The problem of psychomotor retardation among dysthymics was discussed briefly.

6. It was predicted that a certain type of distraction would temporarily obscure or break up the pattern of affective disturbance and would result in a speeding up of

performance on the Maze test in the case of dysthymics, but no such speeding up in the case of non-dysthymics. Both these predictions were upheld by the experimental findings.

7. It was predicted that after electroconvulsive therapy dysthymics would give a score pattern more akin to that of the non-dysthymics. This was borne out by the experimental findings. The effect on Maze performance of distraction and E.C.T. was, therefore, similar. An explanation in terms of partial deflection of attention was suggested.

8. It was claimed that the effect of distraction on speed of performance and certain other scoring categories on the Maze test is a useful method of helping to assess the intensity of an affective disturbance.

References

(1) Adrian, E. D. (1944). Brain rhythms. *Nature, Lond.*, cliii, 360.

(2) Cameron, N. (1942). The place of mania among the depressions from a biological standpoint. *J. Psychol.* xiv, 181–95.

(3) Downey, J. E. (1924). *The Will-Temperament and its Testing.* New York.

(4) Foulds, G. A. (1951). Temperamental differences in Maze performance: Part I. Characteristic differences among psychoneurotics. *Brit. J. Psychol.* xlii, pt. 3, 209–17.

(5) Hovey, H. B. (1928). The effects of general distraction on the higher thought processes. *Amer. J. Psychol.* xl, 585–91.

(6) Laird, D. (1929). Experiments on the physiological cost of noise. *J. Nat. Inst. Industr. Psychol.* iv, 5.

(7) MacCurdy, J. T. (1925). *The Psychology of Emotion.* London.

(8) McFarland, R. A. & Goldstein, H. (1939). The biochemistry of manic-depressive psychosis. *Amer. J. Psychiat.* xlvi, 21–58.

(9) Morgan, J. J. B. (1916). Overcoming of distractions and other resistances. *Arch. Psychol.* v, 35.

(10) Sargant, W. (1951). The mechanism of 'Conversion'. *Brit. Med. J.* 11 August.

(11) Super, D. S., Braasch, W. F. & Shay, J. B. (1947). The effect of distraction on test results. *J. Educ. Psychol.* xxxviii, 373–7.

(12) Weber, H. (1929). Untersuchungen über die Ablenkung der Aufmerksamkeit. *Arch. ges. Psychol.* lxxi, 185–260.

(13) Wiener, N. (1948). *Cybernetics.* New York.

(14) Wolff, H. G. & Goodell, H. (1943). The relation of attitude and suggestion to the perception of and reaction to pain. *Pain, Res. Publ. Ass. Nerv. Ment. Dis.* xxiii. Baltimore.

(*Manuscript received* 8 *February* 1951)

THE CLASSIFICATION OF MOTOR PERSEVERATION TESTS: SOME CRITICISMS AND DEDUCTIONS

By E. PRATT YULE

The Department of Psychology, University of Natal

I. *Theories of p-factor* (p. 42). II. *Classification of p-tests* (pp. 43–44). III. *Criticisms of classificatory criteria* (pp. 44–46). IV. *Non-validity of classification* (pp. 46–47). V. *Deductions and implications:* (i) *Failure to test the p-hypothesis* (pp. 47–48); (ii) *Confusion of abilities and traits* (pp. 48–49); (iii) *Advantages and limitations of projective tests* (pp. 49–50); (iv) *Failure of objective quantitative tests of personality* (pp. 50–51); (v) *Clinical situations as a source of information about personality* (pp. 51–52). VI. *Summary* (p. 52). *References* (p. 52).

I. Theories of *p*-factor

As a psychological artefact, the concept of perseveration has enjoyed a surprisingly long career. After a quarter of a century of description and exploratory measurement it was received into the factorial fold with the publication of Spearman's(15) *Abilities of Man* in 1926. In the ensuing twenty years there were many investigations of *p*-factor, and considerable discussion of its real existence and nature. In the earlier years of this period, up till about 1935, the theory and practice of *p*-testing was in a state of confusion, but within the next decade the position became clearer. It appeared that *p*-factor could be obtained only from motor tests. Two distinct concepts of its nature crystallized out of the confusion of theories and equivocal experimental results. *p*-Tests were subsequently classified into two types according to which hypothesis of the nature of *p*-factor they were designed to investigate.

The first theory in the field was that of psycho-physical lag, secondary function, (Wiersma(24)) or 'mental inertia'. It explained *p*-scores in terms of the effects of one discrete sensory-motor process on a consequent one. The implication of this view is that such processes can be considered in isolation from the total behavioural continuum. Such a view reflects the analytical atomistic approach of academic psychology at the turn of the century, and it is not surprising that it had lost ground by 1930. Stephenson(17) accepted it as only one possible interpretation of *p*-factor. In view of the claims that *p*-scores correlated highly with qualities of character(9, 2, 12), he suggested that *p*-scores might directly or indirectly measure *W* characteristics(23), such as effort in the face of difficulty. Cattell(4) extended the general trend of this suggestion into a theory of 'disposition rigidity'. All such views interpret *p*-scores in terms of some sort of ego function, thus implying the operation of the whole structure of personality.

By 1935 the two interpretations of *p*-factor had been clearly enough distinguished by Cattell, and on the basis of the distinction *p*-tests used in some later investigations were classified into 'creative effort' and 'alternating' types, with different content, arrangement, timing and scoring. It is with this classification, as it appears in the work of Notcutt(11) and Cattell(3, 4), that this discussion is concerned.

II. Classification of p-tests

Creative effort tests are presumed to measure the subject's reactions to the breaking of a familiar habit, such as writing 'it' in the usual way (the X-activity), and then with backward strokes (the Y-activity). p-Score is the ratio of X-activity to Y-activity measured by number achieved in unit time. It is assumed to represent some ego variable.

Alternation tests are designed to measure the after-effect of the first, or X-activity, on the second, or Y-activity, when they follow each other in rapid temporal succession. Thus the tests are constructed so that a subject performs two activities independently and then in quick alternation. p-Score, on the scoring system devised by Notcutt ((11), p. 202), is measured by the difference between the predicted and the actual number of responses achieved in the alternating activity. It is assumed to represent the inertia of psycho-physiological processes evidenced in the ease with which the subject can switch from one process to another.

What basic differences are being presumed to exist between these two types of p-tests? Both demand the juxtaposition of two activities, X and Y. But since each is designed to test a different hypothesis about the nature of p-factor, the types differ in respect of: (a) arrangement, timing and scoring; (b) the nature of the activities involved and the task set to the subject. Under (b) there are two aspects to be considered. There is: (i) the relation of the visual end-product of the Y-activity to that of the X-activity; and (ii) there is the nature of the sensory-motor demand made on the subject. For example, if the Y-activity of a test demands that the subject write the word 'it' backwards, the visual end-product of this activity should correspond to that of the X-activity. If he is asked to write 'S' as it would appear in a mirror, then the visual end-product of the Y-activity is different from that of the X-activity. As for the nature of the sensory-motor demand made on the subject, the question whether the task set in writing 'it' backwards is more or less difficult than writing 'S' mirror-wise, has to be considered.

It would be premature to consider the arrangement, timing and scoring of the tests before we have considered their content, for the former depends on the latter. In order to get some light on the difference in the content of the two types of tests, the views of Notcutt (11) and Cattell (4) will be examined. We are told that in creative effort tests, the X-activity should consist of the performance of some old, accustomed task. The Y-activity should demand its performance in another way which would involve the breaking of the habit. This should not, however, demand intelligence. Thus, to the Y-activity the words 'new', 'unfamiliar', or 'abnormal' are applied. Even if it must not be difficult enough to involve G-factor, it must present some difficulty. Stephenson (17) said that significant p-saturation would only be obtained if the Y-activity was somehow more difficult than the X. Cattell (4) mentions that the Y-activity should be 'similar' in nature and equal in 'complexity' to the X.

It is necessary to add a special note on the work of Walker, Staines & Kenna (21, 22). As a result of their critical examination of methods of constructing and scoring motor p-tests, these investigators reached conclusions about the classification of the tests very similar to our own. They also, however, impute to the X- and Y-activities, qualities such as 'difficulty' and 'newness' ((22), p. 59). Eysenck (6) characterizes the X-activity as 'customary', 'straightforward', and the Y-activity as 'unusual', 'inverted'.

In alteration tests, the activities must be similar, equivalent and comparable in habituation and difficulty.

Clearly, the differentiation between these two types of *p*-tests depends on the characteristics of the *X*- and *Y*-activities and the relation of one to the other. The description of these activities given by the investigators turns out, on analysis, to be couched in terms of the same variables for both types of test. The variables are: familiarity, novelty, difficulty, complexity, similarity, equivalence, normality. (This last is mentioned on two occasions by Notcutt ([11]), p. 202, 204). He appears to mean that the *Y*-activity should demand an 'abnormal' way of carrying out the task set in the *X*-activity, but the point is further discussed below.)

III. Criticisms of classificatory criteria

The most cursory consideration shows that the tests are being classified on the basis of qualities attributed to the *X*- and *Y*-activities, and that these qualities are being presumed to be absolute. But quality judgements are especially subjective and can only attain a measure of objectivity if they yield a high percentage of agreement over a large number of judgements. Complete agreement is rarely reached, a fact adequately demonstrated by the limitations of nineteenth-century psycho-physics. The writer has been unable to find, in the literature dealing with *p*-tests, any record of a preliminary investigation into the range of judgements about the nature of the tasks presented in the tests. Thus the description of the two activities in terms of the variables catalogued above has no objective validity. It represents the subjective judgement of those constructing the tests. The true nature of the demands made by any *p*-tests remains a function of the way in which the subjects experience the *X*- and *Y*-activities and their juxtaposition. Thus, *p*-score must be influenced by subjectively determined perceptions for which the investigators have made no allowance.

We shall examine a few instances where the investigator's attribution to the *Y*-activity of supposedly objective qualities such as 'familiarity' 'difficulty' or 'abnormality' breaks down. The H I test used by Notcutt and Cattell is classified as 'creative effort'. If it is to qualify as such, the symbol I must be unfamiliar to all subjects. It would not, however, be unfamiliar to any subjects who had some acquaintance with mathematical symbols. In one of our own investigations θ was employed as a unit. In our university group there must have been subjects to whom this was genuinely unfamiliar, but a few to whom it was a known Greek letter. This must apply to many of the symbols which have been used in *p*-tests with adult—especially university or professional—groups ([11, 3, 4, 12]). Again, in the I test, one subject may see the I as 'familiar-letter-H-lying-on-its-side', while another may regard it as a totally unfamiliar gestalt.

Another notable factor entering all *p*-tests is that speed is mentioned in the instructions given by most investigators. In an investigation conducted by the writer ([26]) subjects were requested to write with 'as much speed as was compatible with reasonably well-formed letters and figures'. Here again, of course, subjects interpret these instructions in terms of subjective standards. A point of interest is, however, raised by the frequent occurrence of imperfectly formed letters or digits. These clearly present many possibilities for the formation of new gestalten. Suppose, for example, that in the I test, the subject's I looks rather like an I because of poor formation. That subject may then treat an alternation of H I as a gestalt composed of the alternating letters H I, H I. Again, the writer has observed the mirror image of the letter S become, as a result of hasty writing, an imperfect

igit '2'. The *Y*-activity of this subject was thus simply the writing of 'S 2', 'S 2', 'S 2'. -Factor cannot be supposed to enter either of these tasks.

Although the form in which the instructions are given to the subject will certainly ıfluence gestalt formation, this can only be determined within certain broad limits, and ıe range of deviation must be considerable. All the findings concerning the influence of ıbjective attitudes on gestalt formation are relevant here. The problem resembles that ıced by the early investigators of memory when they tried to draw up lists of meaningless r nonsense syllables. Meaning, familiarity, difficulty, complexity are qualities of gestal-ɜn, so also are the qualities of similarity, equivalence and normality mentioned by lotcutt and Cattell in special reference to the activities required in alternation tests. All ıch qualities lie in the eye of the beholder, and who but the subject shall say whether, in test where the *X*-activity is $\pm$ and the *Y*-activity is ∞, these are even perceptually quivalent? To a subject who knows mathematics they may be equally familiar. A school-ıild might regard them differently. The question of the equivalence of the motor task ıvolved in their actual formation has also to be considered.

The terms similarity, equivalence and normality, as used by Cattell and Notcutt, of the ctivities in alternation tests involve some ambiguity. Sometimes they are used to apply ɔ the way of performing an activity, i.e. to the motor side of the performance, and, at ther times, to the visual end-product of the activity. (It was for this reason that the riter, in § II, distinguished the activity itself from its visual end-product.) For example, otcutt uses the term 'normal' to apply to the test where '2 3 4' is written normally and ıen with reverse strokes. This is classified as a creative effort test along with the H I-test. ut these two tests are not comparable. In the '2 3 4'-test the visual end-product of the -activity should be the same as that of the *X*-activity, and is in no sense new or unfami-ar. But in the H I-test the product of the *Y*-activity is a different perceptual gestalt om that resulting from the *X*-activity. Further, the motor operations involved in roducing '2 3 4' by reverse strokes do not appear comparable or equivalent to those ıvolved in producing H I.

Here are two tests classified together, yet they do not appear structurally equivalent, uite apart from the way the subjects may regard them. Throughout the history of the -testing there appears to have been neither detailed analysis of the demands made by the ɜsts on the subjects, nor scrutiny of the structural equivalence of the tests.

The whole concept of a 'normal' way of forming symbols illustrates the point at issue. here probably is a normal way of forming the letters, digits and symbols which are the ɔmmon material of paper-and-pencil motor *p*-tests. Methods of teaching children these hings are relatively standardized. But the concept of normality includes the range of eviations from the norm, and such deviations must be significant when the comparability f the tasks is being considered. The writer has observed school children, both in life tuations and in the *X*-activities of *p*-tests, writing letters and figures in ways deviating ɔnsiderably from the observed norm. Such deviations must affect the subject's attitude ɔ the activity demanded, and hence his performance on both *X* and *Y*-activities. uppose one subject habitually forms the digit '2' with reverse strokes (and the writer as observed this happening) then the *Y*-activity in a '2 2' test is the same as the *X*, and ıe *p*-scores of this subject in this test represents nothing but chance variations in speed.

Finally, we have considered another activity which was frequently used by investi-ıators. Subjects were asked to write familiar letters or digits as they would appear in

a mirror image. Wynn Jones (16) was using such a test in 1915, and batteries used more recently by the writer and others have sometimes demanded *Y*-activities involving mirror-image writing, and sometimes backward-stroke writing. For example, one common test is 'S' followed by 'Ƨ'. Another is '*b*' followed by '*b*' written with reverse strokes. This raises another aspect of the problem of differential gestalt formation. If rapid alternation is required, it is clear that 'S Ƨ' may rapidly be adapted to a new visual and kinaesthetic gestalt. In such cases, the alternating performance of 'S Ƨ' is not really comparable to that of writing '*b*' with backward strokes. The point at issue is not only that the visual end-products of the two activities differ, but the possibility that 'S Ƨ' lends itself to gestalt formation, both visual and kinaesthetic in a way which '*bd*' does not. The ability to organize perceptual data into fresh gestalten may or may not have something to do with *p*-scores, but what is significant is that there does not seem to be any reference to this possibility in the literature.

Another aspect of tests demanding the formation of mirror images should be critically considered. This is the possibility that such tests may measure an unknown quantity of *K*-factor (space relation ability). Describing the essential nature of this ability, Stephenson (18) says 'high *K* is shown by the child who can draw the letter *R* as it would appear in reflexion in a pool of water without having to think explicitly about it'. It would be reasonable to suppose that tests demanding mirror image formations would, in E. Koussy's (5) words, involve 'the ability to obtain and the facility to utilize visual spatial imagery'. If so, tests in which the *Y*-activity is of this nature may have quite a high degree of *K*-saturation. Mirror writing is used in the *Y*-activity of *p*-tests in the belief that it presents a task of equal difficulty to all subjects and that what *p*-score measures is the subject's varying reactions to this standard task. But according to the findings on the nature of *K*-factor (14, 1, 20) the *K*-saturation will vary with the age and sex of the subjects. Thus, if mirror-image *p*-tests do involve this factor, the assumption that they present a task of equal difficulty to all subjects is clearly invalid. Subjects of different age, sex and level of *K*-ability will regard the *Y*-activity differently, and the *p*-score will be correspondingly contaminated.

Our remarks have been confined to tests involving mirror imaging. Many tests with high *K*-saturation, however, involve imaging and/or rotating a figure into another position. This may be precisely what is demanded in the 'H'-test. Finally, there is the fainter possibility that writing with reverse strokes may also demand something of this 'facility to utilize visual spatial imagery'.

IV. Non-validity of classification

Two conclusions can be drawn from these considerations:

(i) *The heterogeneity of p-test batteries.* Critical examination of the activities demanded in *p*-tests leads to the conclusion that *p*-tests classified as of one type may set tasks to the subjects which are not perceptually equivalent, as for example, in one test, the visual end-product of the *X*-activity is like that of the *Y*, while it is different in another test. The existence of such anomalies means that any one supposed type of *p*-test battery may turn out on examination to contain a structurally heterogeneous collection of subtests making unanalysed demands on the performers. Such differences have lain concealed under portmanteau terms like 'interference', 'inertia', 'effort'. This at least could partially account, in a battery of subtests, for the consistently low intercorrelations between

subtests which the investigators rated as similar in design and demand on the subjects. It is of little use to argue that from such a random sampling of reactions to juxtaposed activities, factor analysis must produce a *p*-factor as it once so obligingly produced a *G*-factor. The situation is quite different. The nature of the difference is discussed in § V, 2.

(ii) *Variation in the test situation.* No investigator can tell in advance how the tasks set in *p*-tests will appear to his subjects. The nature of the demand will vary from subject to subject as the personality pattern varies. *p*-tests have attempted to deduce personality variables from measurements of reactions to situations which are presumed to present standard problems to subjects. In fact, since the nature of the problem is a function of individual perception, the situations can theoretically present a range of variation as great as the number of subjects being tested.

Recent investigators have recognized that those processes to which the term 'perseverative' was applied do not belong to a single functionally unitary tendency (4), and that *p*-score does not measure a unitary factor. It is affected by many variables, some of which have been pointed out by older investigators. The influence of *G*-and *F*-factors has been considered. The writer has suggested the possible intervention of *K*-factor in some types of subtests. Consideration of speed was reflected in improved methods of scoring, and most investigators have estimated the effects of age and sex: but no investigator, to our knowledge, has considered the most fundamental variable of all—that which arises from variations in the meaning of the test situation itself. If the problem has been considered, it has not been met, for the methodological answer would be a factorial experiment using the techniques of analysis of variance. Of this there is no record, and, in any case, use of these methods on data obtained from current forms and classifications of *p*-tests would be valueless for reasons now to be examined.

If the meaning of the test situation is a variable, and no objective characteristics can be imputed to the *X*- or *Y*-activities in *p*-tests, then *p*-tests cannot be classified into creative effort and alternation types in terms of the nature of the activities involved. The only remaining distinction between the two tests lies in the arrangement and timing of the juxtaposed activities: but without any valid distinction between the kinds of task set, any necessity for this falls away. For example, if there is no guarantee that the *Y*-activity in a creative test demands the formation of an unfamiliar visual symbol, or involves difficult or 'abnormal' motor reactions for all the subjects being tested, then there is no point in arranging, timing and scoring the tests of the creative type differently from those of the alternating type.

In short, the classification of *p*-tests into creative and alternating types breaks down not only with regard to internal structural consistency, but also as a qualitatively exclusive classification.

V. Deductions and implications

(i) *Failure to test the p-hypothesis*

If we are justified in our view that the classification of the tests is invalid (and it is the only clear classification to be found in the literature) other deductions can be drawn. The differentiation between *p*-tests arose in the first place because each type was designed to test a different hypothesis about the nature of *p*-factor, i.e. whether it represented the operation of psycho-physical or characterological factors. If the differential criteria of

the two types of test break down in practice, there is no guarantee that any one type of test presents to the subjects the type of situation the experimenter believed it to present. There is also no guarantee that experimenters have ever measured what they believed they were measuring. Hence it seems open to doubt whether: (i) either hypothesis of the nature of *p*-factor has ever been tested, and hence whether (ii) the existence of perseveration has been either proved or disproved.

There are some facts about the concept of perseveration which should not be lightly passed over. The long history of attempts to measure '*p*' indicates a persistent belief in a cluster of variables describable under the broad general concept of perseveration. Psychological literature is full of descriptions of personality differences couched in some such terms. For example, Freud(7) describes people who 'cannot...detach libidinal cathexes from one object and displace them to another...and the opposite type in which the libido seems especially mobile and readily enters new cathexes'. Cattell's discussion of the possible correlates of 'disposition rigidity' points to the same fact, i.e. that a lot of different aspects of behaviour seem roughly related to the broad general concept of perseveration. Belief in such a concept cannot be based on experimental evidence, for this has been unsatisfactory. It must rest on one of those intuitive judgements derived from observation of total behaviour in life or clinical situations. Such judgements should not be lightly dismissed, for they are sources of testable hypotheses about the nature of behaviour. Many of the hypotheses of the psycho-analysts are of this nature. The belief that there is something in the concept of perseveration is, we think, of this type. It may be right or wrong, but the significant fact is that thirty years of experimentation with *p*-tests have not been able to prove or disprove the hypothesis. Interest in *p*-factor and *p*-tests may now have waned. Nevertheless, the failure of the whole line of thought centered round *p*-testing has implications for the field of personality measurement in general, and reflects on certain central contemporary issues.

(ii) *Confusion of abilities and traits*

The failure to test the *p*-hypothesis was the result of unconscious assumptions made by those who first set out to measure perseveration. Only one area of behaviour has so far proved amenable to direct quantitative measurement. This is the relatively narrow field of operations falling under the head of test intelligence. Before intelligence was measured at all, there was a consensus of agreement about the characteristics of intelligent behaviour and the sort of situation in which it was elicited. The first rough samplings of it in test situations gave promise of consistency and validity, and by the first decade of this century it was clear that certain abilities could be quantitatively estimated. It was this success with intelligence which seduced psychologists into drawing analogies between abilities and personality traits, and supposing that the same procedures were applicable to both. Early investigators of perseveration clearly believed that if measures of performance in a large number of test situations involving juxtaposed activities were pooled, *p*-factor could be extracted just as '*G*' had been extracted (see(16), pp. 156, 157 and 159).

The situation is quite different for personality traits, and the differences should be clearly recognized. The factor for general intelligence emerges from batteries of tests which, on the last analysis, measure only one ultimate variable, the adequacy of the subject's perceptions of external reality, or, in normative terms, their truth or falsity. (Special abilities refer to special ways of dealing with these perceptions, as, for example,

p-factor refers to the ability to manipulate external reality in terms of verbal symbols.) The field of behaviour described as intelligent is delimited by reference to this variable. The situations involved are those in which adaptive behaviour depends on adequate perception of the external world. Intelligence tests present artificial situations of this type, and because of the unequivocal nature of the criterion, such situations are not difficult to construct, to delimit or to evaluate in terms of the criterion of perceptual adequacy. In short, intelligent behaviour occurs in one delimited type of situation where the truth or falsity of perceptions is involved. Behaviour in such situations is either intelligent or not intelligent. Intelligent behaviour is not the mode of a range of responses. But no one ultimate variable delimits any field of that behaviour whose totality is called personality. Personality traits refer to an individual's reactions to situations. No two reactions of an individual are ever precisely alike; the situations to which the individual responds are never precisely the same. A personality trait is thus the mode of a range of behavioural reactions to a modal situation. The behaviour in question is not, except indirectly, related to the adequacy of the individual's perceptions of the external world. (Social psychologists will see the distinction as that drawn by such writers as Kardiner and Linton between reality and projective systems. Psycho-analysts will relate it to the distinction between Freud's pleasure and reality principles.)

Since the behaviour falling under the head of personality cannot be delimited as intelligent behaviour can, the situations which elicit it cannot be neatly circumscribed. It thus becomes very difficult indeed to construct artificial or test situations which will present the same essential features. Because the situations which would elicit behaviour having a certain quality such as for example, aggressiveness, cannot be clearly delimited, they will not resemble in the least the sort of situation which would elicit behaviour having the quality of intelligence. Investigators of perseveration did not realize this. They hoped that by presenting the individual with a few situations involving juxtaposed activities with a pencil and paper, they would deduce a behavioural character running through a multiplicity of life situations and give it a quantitative value. The ultimate cause of their failure to do so lay in their failure to distinguish between abilities and personality traits and the nature of the situations involved.

(iii) *Advantages and limitations of projective tests*

The failure of all tests constructed on the same assumptions as *p*-tests to give any significant information about personality was no doubt responsible in part for the enthusiastic reception of projective tests like the Rorschach and T.A.T. These grew out of a different tradition of thinking from that which produced the investigation of abilities and personality factors like '*W*', '*f*' and '*p*'. They have corrected the major faults of the personality factor type of test. They involve no attempt to delimit the test situation which, as we have seen, cannot be delimited as can the intelligence test situation. In so far as the situation in projective tests involves external stimuli at all, these are secondary to the meaning imposed by the subject on them. The definition of the test situation becomes a matter of the meaning the subject reads into it, and the very variations in meaning become the data on which deductions as to personality are based. The impossibility of delimiting the test situation which was a drawback to tests like *p*-tests becomes a virtue in projective tests.

Projective techniques have enjoyed more than a decade of widespread support, but rifts

in the lute are appearing. It may be that such tests give no more information about a individual personality than a skilled interviewer would obtain in the same or less time Yet, since the tests cannot be interpreted by anyone who has not the same knowledge o human behaviour as a skilled interviewer, they are neither a vital addition to his equip ment, nor do they save his services. Our hypothesis is that all the faults inherent in th older type of personality tests have not been overcome by projective techniques. Althoug they take larger, longer, relatively undetermined samples of individual behaviour, th samples are still infinitesimal parts of the behavioural continuum. 'The T.A.T. draw forth no more than twenty small samples of the subject's thought. To suppose these wil invariably provide a skeleton of total personality is unduly optimistic' (10). The man people who suppose that they will are simply repeating—although on a much smalle scale—the faults of the personality factor testers.

And how, in the end, are these samples of behaviour interpreted? The reference i always to inductions based on long-term observation of total behaviour. The ver construction of projective tests is based on these. Rorschach was, above all, a psycho analyst, and his test was designed to show the operation of Freudian dynamics. Murra stressed the dependence of his system of T.A.T. interpretation on the psycho-analyti theory of personality. Korner (8), reviewing the limitations of projective tests, points ou that interpretations are made through inferences depending on the interpreter's knowledg of the dynamics of personality. Rosenzweig (13) says that projective tests represent i considerable degree the premature application of just emerging psychodynamic principles The tests cannot, in fact, go beyond the interpreter's already existing knowledge o personality. Most of this knowledge will have been derived from observations mad in clinical practice. Despite this fact, some psychologists interpret data from Rorschac tests by rule of thumb and imagine they thereby attain objectivity. The fallacy i obvious; no test can be more objective than the assumptions and principles upon whic it is based. This fact holds for all types of personality tests, including those which clain to be objective and quantitative.

(iv) *Failure of objective quantitative tests of personality*

We have examined some aspects of *p*-tests in detail. But the *p*-test is only one repre sentative of a type of test which has been described as objective and quantitative, and o a certain line of thinking about the nature and measurement of personality. The failure o *p*-tests is symptomatic of the failure of this line of thought. Consideration of two relativel recent publications confirms this view. In *Human Ability* (16), the summary of research o abilities carries conviction; that on personality traits does not. It is interesting that th two are scarcely differentiated; *p*-factor is discussed as if it were not markedly differen from '*G*'-, '*V*'- or '*K*'-factors, except through an accident of historical origin (p. 165).

Dimensions of Personality (6) claims to be a 'rigorously objective investigation dealin with operationally defined concepts throughout' (p. 16). The investigations reporte naturally involve specific assumptions about the methods by which personality must b investigated: 'any view of personality must be based on experimental results treated b statistical methods'. Sutherland (19) has criticized this whole school of thought for hopin to identify what is dynamically relevant by purely experimental and statistical analysis Our criticism approaches the matter from another, but related, angle. Personality test such as those for persistence, rigidity, fluency, suggestibility, place the subject in highl specific artificial situations of brief duration. The life situations in which personalit

raits manifest themselves are not of this nature at all. Thus, behind the experimental et-up lie a number of unstated hypotheses about the nature of personality. First, there s the assumption that personality traits can be treated in the same way as abilities, which s probably based on failure to distinguish between the two. Again, there is the assumption that samples of behaviour obtained in such situations can validly represent the total nterdynamic continuum of individual behaviour. On this assumption, the quantitative lata obtained from these samples are subjected to elaborate statistical analysis, out of vhich factors said to represent behavioural modes are drawn.

It is evident that implicit assumptions about the nature of personality, and the way in vhich it manifests itself, determine the situations from which the data about personality re collected. However careful the statistical analysis of such data, the validity of the onclusions drawn from them cannot be greater than the validity of the assumptions vhich determine the experimental set-up, i.e. in this case, the nature of the test situation.

These assumptions, however, do not appear more valid or objective than the assumptions lying behind the collection of data from projective tests. For reasons stated in ection V, ii and iii they are less so.

(v) *Clinical situations as a source of information about personality*

It is admitted that projective tests depend on the interpreter's existing knowledge of ehaviour, and that the reference is back to inductions based on long-term observations f persons, most of which have been made in clinical situations. It is less generally ecognized that so-called objective tests of personality also refer back to this type of udgement. The very construction of such tests involves assumptions about the nature of ersonality not themselves arising from experiment. Further, throughout their history, common method of validating tests of fluency, perseveration, persistence, etc., has een to apply them to individuals whose personality structure had been clinically escribed (25, 9, 12, 6). The observations which in the past, and still, are the source of uitful hypotheses about the nature of personality have come out of clinical situations. o conventional type of personality test has gone beyond the knowledge already possessed s a result of the accumulation of clinical observations.

It is a commonplace that experiment comes after hypothesis and that every science oes through an anecdotal stage. It is often argued that inductions from clinical observaon represent this stage, and that psychologists must step in to validate the often merely ntuitive judgements arising out of clinical experience. We may certainly agree that this the task of psychology; it is the methods proposed which should be critically examined. he construction of limited test situations which are presumed to isolate some personality nction for quantitative assessment and statistical analysis is a method doomed to ilure. The expression of personality cannot be narrowed down as can the expression of ntelligence. The history of attempts to measure personality demonstrates that no test tuations so far devised are superior to clinical situations as a source of information about ehaviour. Observation in the clinical situation does not merely represent a pre-scientific age of observation. Other sciences have passed from the stage of hypotheses to that of xperiment, and found in experiment the main source of new hypotheses. So far, this is ot true of the experimental approach to personality, objective or projective. It has roduced some verifications of clinically founded hypotheses, but it has not gone beyond ese hypotheses to produce new ones. Thus it is that in the end we turn to clinical bservation, and it is the clinical situation which should become the prototype of situa-

tions for the scientific observation of behaviour. Instead of encouraging the mushrooming of tests, psychologists might more wisely turn to devising methods by which data obtained in clinical situations could be used to reach scientifically valid conclusions, while preserving all that is most valuable in the clinical situation. Such would be the relation of person with person, and the careful, prolonged and detailed study of persons.

VI. Summary

A conventional classification of motor *p*-tests is examined and rejected. The conclusion is drawn that the existence of *p*-factor has been neither proved nor disproved despite over thirty years of experiment. It is argued that *p*-tests and all similar so-called objective quantitative tests are based on invalid assumptions, and notably that they involve a failure to distinguish between abilities and personality traits. Projective tests are superior in this respect, but all tests of personality—objective or projective—rest on assumptions largely derived from observations of behaviour in clinical situations. Such situations remain the best source of data on personality, and should become the prototype of situations for the scientific observation of behaviour.

References

(1) Adcock, C. (1948) A re-analysis of Slater's spatial judgement research. *Occup. Psychol.* xxii, no. 4 1949.
(2) Bernstein, S. (1924). Quickness and Intelligence. *Brit. J. Psychol.* (Monogr. Suppl.), vii, 49.
(3) Cattell, R. B. (1935). On the measurement of perseveration. *Brit. J. Educ. Psychol.* v, 76.
(4) Cattell, R. B. (1946). (1) The riddle of perseveration. (2) Solution in terms of personality structure. *J. Personal.* xiv, 239–67.
(5) El Koussy, A. (1935). Visual perception of space. *Brit. J. Psychol.* (Monogr. Suppl.), no. 20.
(6) Eysenck, H. J. (1947). *Dimensions of Personality*, p. 157. London: Kegan Paul.
(7) Freud, S. (1937). *Collected Papers*, v, 344. London: Hogate Press.
(8) Korner, A. F. & F. (1950). Theoretical considerations concerning the scope and limitations of projective techniques. *J. Abnorm.* (*Soc.*) *Psychol.* xlv, no. 4, 619–27.
(9) Lankes, W. (1915). Perseveration. *Brit. J. Psychol.* vii, 387.
(10) Murray, H. A. (1943). *Manual: Thematic Apperception Test.* Harvard University Press.
(11) Notcutt, B. (1942). Perseveration and fluency. *Brit. J. Psychol.* xxxiii, 200–8.
(12) Pinard, J. (1932). Tests of Perseveration. *Brit. J. Psychol.* xxiii, no. 5, 114.
(13) Rosenzweig, S. (1948). *Orthopsychiatry* 1923–1951, pp. 456–66. Amer. Orthopsych. Ass. Inc.
(14) Slater, P. (1940). Some group tests of spatial judgement. *Occup. Psychol.* xiv, 45.
(15) Spearman, C. (1926). *The Abilities of Man.* London: Macmillan.
(16) Spearman, C. & Wynn Jones, Ll. (1950). *Human Ability.* London: Macmillan.
(17) Stephenson, W. (1934). An Introduction to so-called Motor Perseveration Tests. *Brit. J. Educ. Psychol.* iv, 186–208.
(18) Stephenson, W. (1950). *Testing Schoolchildren.* London: Longmans, Green and Co.
(19) Sutherland, J. D. (1950). *Universities Quarterly*, iv, no. 2, 157–67.
(20) Swineford, F. (1948). The nature of the general verbal and spatial bi-factors. *Univ. Chicago Suppl. Ed. Monogr.* no. 67.
(21) Walker, K. F., Staines, R. G. & Kenna, J. C. (1941). Is there a General Factor of Perseveration? *Aust. J. Psychol. Phil.* xix, 58–75.
(22) Walker, K. F., Staines, R. G. & Kenna, J. C. (1945). The influence of scoring methods on score in motor perseveration. *Brit. J. Psychol.* xxxv, no. 3, 51–60.
(23) Webb, E. (1915). Character and Intelligence. *Brit. J. Psychol.* (Monogr. Suppl.). No. 3.
(24) Wiersma, E. (1906). Die Sekundarfunktionen bei Psychosen. *J. Psychol. Neurol., Lpz.*, viii.
(25) Yule, E. Pratt (1935). The resemblance of twins with regard to perseveration. *J. ment. sci.* lxxxi, 489–501.

(*Manuscript received* 9 *July* 1951)

FAMILY PICTURES: A PROJECTION TEST FOR CHILDREN

By JEAN D. CUMMINGS

Introduction (p. 53). II. *Test procedure* (pp. 53–54). III. *Illustrative child guidance cases* p. 54–56). IV. *Interpretation of test results* (pp. 56–60). V. *Summary and conclusions* (p. 60).

I. Introduction

hatever one's approach to children's difficulties, whether it is from a remedial teacher's ewpoint, or that of a psychologist at a diagnostic interview, or of a play therapist, common need is to understand the child's attitude to his home and parents; this must clude both his conscious feelings and also the phantasies of parental anger and punish- ent that children experience. It is also of practical importance to know the kind of me situation that produces feelings of guilt and anxiety. For many children, particularly ung children and those of subnormal intelligence, verbal expression is difficult. For hers the direct expression of hostility towards their parents is provocative of guilt and erefore inhibited. It is well recognized that such feelings can have indirect expression dreams, and they can also have expression in a projected form in material where the ild can describe a boy or girl and the parents without having necessarily to acknowledge at he is describing himself. It is not possible to disentangle how much the child projects own actual situation and how much is phantasy, but in either case a good deal of his n attitude is bound to be revealed.

As a projection test convenient for use with young children the Family Picture test has oved to be promising. It has been tried out by the author with boys and girls attending Child Guidance Clinic and with a control group. The test and some results are here scribed, as other psychologists may be interested in the test as a tool.

II. Test procedure

e test takes about 10 min. to give, and is designed especially for children whose verbal pression is limited. Young children of 5 to 6 particularly enjoy being asked to make up ries about pictures, and like to give full descriptive details.

The first picture to be shown is of a child of the same sex as the one being tested. The ial expression of the child is as neutral as possible. The boy is asked simply 'what shall call him?' and then 'what is he like—tell me about him'. This usually elicits a simple scription, often of the child's hair or clothes. In the second picture, the parent of posite sex is shown, again with a neutral expression, and comments are invited. llowing this there is a picture of the other parent. At the next stage the parents are own again in turn, but this time they are frowning and the child is asked to say why, at has happened. If he gives some partial explanation, such as that 'something's ppened' he is asked to explain in more detail, or the interviewer simply responds with inquiring 'Oh?'. No emotion in the parents is suggested, the child seeing the frowning rents as cross, ill or miserable as the case may be.

After this he is shown a picture of the child crying, and again asked for an explanation l helped to enlarge on this by remarks that are neutral in content though encouraging tone. The next picture shows the child looking frightened (in this case fright has to be

suggested verbally, the artist having found this difficult to depict). The child again gives reasons for the fear.

At the end the child is shown the pictures of the girl and of a baby, both neutral in expression. Finally he is asked two questions: (1) What do you think of ..... (boy's name)? and (2) Is he like you at all?

For girls the test is exactly the same except that they are shown a picture of a girl, the father is presented before the mother, the crying girl is shown instead of the boy and a 'neutral' boy is given at the end.

III. Illustrative child guidance cases

The following three cases illustrate the kind of response given by children of different sex and intelligence level.

Case 1. *T.G.* (boy, aged 7 years 11 months), referred for backwardness, and babyish behaviour. He was of dull intelligence (Terman Merrill I.Q. 77), very lacking in confidence, discouraged at school and with negligible school attainments. His mother and father were also dull and not much concerned with his problem. The projection story probably reflects without much distortion the boy's own family situation and his own apprehensive outlook on life.

Boy. Called John.
What is he like? No reply.
Mummy. Called Mary. She has dark eyes, black hair. John likes Mum best.
Daddy. No reply.
Mother frowning. Father's hitting the children—she feels like she's going to cry. The children 've been naughty—been jumping over the armchairs.
Daddy frowning. Mum's quarrelling with Daddy—cos he hit the children.
Boy crying. Father's hit him—for getting himself dirty in the yard.
Boy frightened. Seen a rat. He's frightened of bulls, skeletons, giants, alligators, tigers, bears.
Girl. Called Mary.
Baby. The best in the family.
What do you think of John? He's scared of everything.
Is he like you in any way? Not at all.

Case 2. *J.B.* (girl, aged 9 years 6 months, Terman Merrill I.Q. 98), referred for refusal to eat. Her school work was 'average' but very slow; her behaviour at school and in the clinical interview was restricted, and her immediate memory inconsistent. The home background was tense, with periodical quarrels breaking out between the parents, the effect of these on the child being very marked. The mother's account was that J. would fight fiercely with her brother aged 4, but at times they would be very loving and go about with their arms round one another. The picture of J.'s hostility to the brother and guilt feelings about this is shown in her description. She actually found some relief in describing the pictures and was surprisingly expressive after her previously extremely defensive manner when asked anything directly about herself. It is interesting that confronted with the picture of the brother she can say nothing—almost as if the resistance here was too strong to be overcome.

Girl. Called Gillian.
What is she like. Blonde hair—a greedy little girl—she eats all the chocolates out of the tin—she has a little brother—name's John—never plays with him—doesn't like him—he starts falling out with her.
Daddy. Wears glasses. He's nice to Gillian—plays a lot of games with her and with little brother as well Gillian only likes him to play with her.
Mummy. Always shouting at Gillian—she has to have the little boy with her or G. would be hitting him

Fig. 1

Fig. 2

Fig. 3

Fig. 4

Fig. 5

Fig. 6

Fig. 7

Fig. 8

Fig. 9

Fig. 10

Fig. 11

addy frowning. Looks very worried—G. acts so naughty—he's always shouting at her—sometimes when he comes to the house he brings G's little brother something—sometimes for G. when she's a good girl.
ummy frowning. Worried about G. too. Worried about G. hitting John and because she's so greedy. If ever she has company there's no chocolates or biscuits.
rl crying. Somebody's given her a good hiding—for hitting the little brother—can't have fun with the others—she has to go to bed.
rl frightened. Thinks she's going to get a good hiding. John's been to tell her mother cos she hit him.
ny. John (no description).
aby. A little fat baby.
hat do you think of G.? A greedy girl.
ke you? No reply.

Case 3. *J.M.* (girl, aged 10 years, Schonell Junior, 111), referred for anxiety symptoms hich included difficulty in getting to sleep, talking in her sleep, vomiting and 'nebulous ars'. She was an only child, living with her parents in very restricted living space. The other described how J. resented any fuss that her father made of her mother, and how e always tried to break in and claim attention to herself. She would tell tales of one rent to the other, apparently trying to divide them. At school she held aloof from other ildren and was easily provoked to tears.

rl. Called Mary.
hat is she like? She's got fair hair, blue eyes, red cheeks, and is wearing a white, blue and grey skirt and in her hair she's got a red ribbon. She's ever so nice.
ddy. He's got brown hair and is wearing glasses. His eyes are brown and he's got on a blue and white striped shirt and black overcoat.
ummy. She's got brown hair and blue eyes—she's not quite so rosy—she's got a Fair-Isle jumper and little skirt and white pinafore—she's been making tarts—she thinks Mary's a good little girl—she always sets the table.
ddy frowning. He's angry with Mummy for not bringing his razor blades—he wants to shave—she forgot his tobacco as well. He often gets cross with Mummy and shouts—shouts at Mary, whenever she's naughty—she's been ever so naughty, she let the rabbits loose. Sometimes she's frightened, she knows she's got to be a good girl. He calls her a 'naughty, little good-for-nothing'.
ummy frowning. Mary's been a naughty girl—been eating the tarts—been pulling the cat's tail.
l crying. She looks untidy—she's been scolded and told she couldn't go to a party.
l frightened. Big boys have been teasing her—she tried to run away from them—she can't get away—they want to smack and tease her—she'd like to hit them back, but she daren't—they just keep teasing her.
by. He's got a little bit of hair—a new bib, a pair of rompers.
at do you think of Mary? She tries to be a good little girl, but her Daddy won't let her. If she tries to help and accidentally does something wrong, he shouts at her—then she doesn't want to help—daren't in case she gets into more trouble.
e you in any way? A bit, but my Daddy's not quite so unkind.

The three cases given are of different intelligence levels, and the third has a far greater ige of expression and imagination than the first—e.g. in the opening descriptions. e first, on the other hand, probably gives a fairly faithful, unembroidered picture of the me situation, possibly a little dramatized here and there. Quite often during the test hild will lapse into the first person, and sometimes he will compare himself with the ld in the pictures—for example, 'I'm not frightened of the dark like what he is!'

t is possible to use the test as a preliminary to discussing the child's own feelings, by nparison with the boy he has just described, but this is usually not possible at any gth in a diagnostic interview; for therapeutic purposes the test could probably be ended so that fresh pictures were available at different sessions with the psychologist.

t is of course not only the fully descriptive responses that are interesting; often as

much is learnt of the child by his silences. One very tense, emotionally restricted child could give no description at all of any of the pictures—it was as if she feared a threat to her conscious attitude of affection for her mother, an attitude which one knew to be precarious in the extreme. Another child could give no description of his father, though he could of his mother. Another child gave a purely descriptive, neutral account—e.g. the daddy is frowning because 'He's not much hair on', the boy is crying because 'he can't see anything'. In such banal replies one may suspect a resistance to the expression of feeling even in its projected form.

It may be that the test is not such as to bring to light the deeper problems of each child, but rather serves to show his more or less conscious attitudes and also to give some inkling of the social situation in which he lives and as he sees it.

IV. Interpretation of test results

Each description of the pictures is in parts unique—e.g. in the description of the child and his parents in the first three pictures, before any suggestion of anger or fear is shown in the facial expressions. With the later pictures, however, certain patterns of response begin to emerge when many tests have been given. One hundred and ten tests have been given. Sixty of these were to children coming to a Child Guidance Clinic, mainly towards the end of a first diagnostic interview. These clinic children were of ages 5 to 12 inclusive. They had been referred for many different reasons and from different sources, which included schools and school medical officers. There were thirty-one boys and twenty-nine girls, with an I.Q. range on the Terman Merrill test of from 54 to 144. A control group of fifty children of similar age range (twenty-six boys and twenty-four girls) was obtained from six Primary schools. Their names were taken at random from class registers. The I.Q.'s of these children were not, in most cases, known.

In the tables that follow the frequency of types of response have been given for the two groups and for boys and girls of both groups together, using Raven's[5] technique for recording both the full individual record of responses as well as responses classified under certain headings. Many responses are, of course, unique—e.g. the little girl who said the father was cross because 'He's looking at a worm—he's mad at the worm—my Daddy buries worms, he doesn't like to see them.' A proportion of responses could, however, be classified when a common feeling was expressed—e.g. of parental anger or punishment.

Differences between groups were tested for statistical significance by the formula $\chi^2=\Sigma\frac{x+m}{m}$ where the observed frequency $=x+m$ and the expected frequency $=m$. A difference was accepted as significant where $P=0{\cdot}05$ or less.

Table 1. *Responses to picture of the mother frowning*

	Clinic group			Control group			Whole group		
	Boys	Girls	Both	Boys	Girls	Both	Boys	Girls	Both
Mother cross with child	14	9	23	10	12	22	24	21	45
Mother sad	9	4	13	3	0	3	12	4	16
Mother worried or anxious	2	4	6	2	0	2	4	4	8
Mother cross with father	1	5	6	2	2	4	3	7	10
Mother ill	2	0	2	3	2	5	5	2	7
Unclassified	3	6	9	6	3	9	9	9	18
No reply, 'don't know'	—	1	1	—	5	5	—	6	6
Totals	31	29	60	26	24	50	57	53	110

More clinic children than control children see the mother as sad. χ^2 8·51; $P<0{\cdot}01$.

Table 2. *Picture of father frowning*

	Clinic group			Control group			Both groups		
	Boys	Girls	Both	Boys	Girls	Both	Boys	Girls	Both
Father cross with child	14	8	22	13	14	27	27	22	49
Father cross for other reasons	3	5	8	2	1	3	5	6	11
Sad or worried	4	3	7	3	2	5	7	5	12
Cross with mother	2	3	5	—	—	—	2	3	5
Unclassified	6	7	13	8	3	11	14	10	24
'Don't know', no reply	2	3	5	—	4	4	2	7	9
Totals	31	29	60	26	24	50	57	53	110

No significant differences between clinic and control groups.

Table 3. *Picture of child crying*

Response	Clinic group			Control group			Both groups		
	Boys	Girls	Both	Boys	Girls	Both	Boys	Girls	Both
(*a*) Punishment responses									
Been smacked or mother cross	2	7	9	3	5	8	5	12	17
Smacked by father or father cross	14	3	17	7	3	10	21	6	27
Smacked by both parents, or both parents cross	5	2	7	3	2	5	8	4	12
(*b*) All responses									
Punishment by one or both parents	21	12	33	13	10	23	—	—	56
Child 'naughty'	2	—	2	1	—	1	3	—	3
Fallen, hurt himself	3	3	6	3	7	10	6	10	16
Separated from mother	—	3	3	1	—	1	1	3	4
Unclassified	5	9	14	8	6	14	13	15	28
No reply, 'don't know'	—	2	2	—	1	1	—	3	3
Totals	31	29	60	26	24	50	57	53	110

Father smacking or cross; sex difference (whole group). $\chi^2 = 9{\cdot}9$; $P = 0{\cdot}01$.
Mother smacking or cross; sex difference (whole group). $\chi^2 = 4{\cdot}0$; $P = 0{\cdot}05$.
Difference between clinic boys and girls for parental anger or punishment, or child feeling naughty. $\chi^2 = 6{\cdot}58$; $P = 0{\cdot}02$.

Table 4. *Picture of child frightened*

Response	Clinic group			Control group			Both groups		
	Boys	Girls	Both	Boys	Girls	Both	Boys	Girls	Both
(*a*) Expressed fear of parents									
Child frightened of father	7	3	10	4	2	6	11	5	16
Frightened of mother	1	7	8	1	—	1	2	7	9
Of both parents	3	—	3	—	—	—	3	—	3
(*b*) All expressed fears									
Fear of one or both parents	11	10	21	5	2	7	16	12	28
Fear of animals	2	7	9	12	14	26	14	21	35
Of being hit by other children	6	1	7	2	—	2	8	1	9
Of burglars	—	1	1	1	3	4	1	4	5
Of ghosts	—	3	3	1	—	1	1	3	4
Separation from parents	3	—	3	—	—	—	3	—	3
Of the dark	2	—	2	—	—	—	2	0	2
Of bad dreams	1	—	1	—	—	—	1	0	1
Individual	5	5	10	4	2	6	9	7	16
'Don't know', no reply	1	2	3	1	3	4	2	5	7
Totals	31	29	60	26	24	50	57	53	110

Fear of one or both parents; more frequent in clinic than control groups. $\chi^2 = 6{\cdot}27$; $P = 0{\cdot}02$.
Fear of father; more frequent in boys than girls. $\chi^2 = 2{\cdot}14$; $P > 0{\cdot}10$ (not significant).
Fear of animals; more frequent in control than clinic group. $\chi^2 = 17{\cdot}22$; $P < 0{\cdot}01$.

Table 5. *Opinion given on the child described*

	Clinic group			Control group			Whole group		
	Boys	Girls	Both	Boys	Girls	Both	Boys	Girls	Both
Unqualified approval	9	12	21	9	12	21	18	24	42
Qualified approval	5	3	8	2	3	5	7	6	13
Disapproval	10	5	15	8	2	10	18	7	25
Unclassified	3	2	5	2	1	3	5	3	8
No reply	4	7	11	5	6	11	9	13	22
Totals	31	29	60	26	24	50	57	53	110

Disapproval more frequent in boys than girls (whole group). $\chi^2 = 5{\cdot}41$; $P = 0{\cdot}02$.

Table 6. *Identification with child in the pictures*

	Clinic group			Control group			Whole group		
	Boys	Girls	Both	Boys	Girls	Both	Boys	Girls	Both
'Yes, like me' etc.	7	11	18	7	5	12	14	16	30
Qualified, 'like me sometimes' etc.	3	3	6	8	3	11	11	6	17
'No, not like me'	16	9	25	7	8	15	23	17	40
'Don't know', no reply	5	6	11	4	8	12	9	14	23
Totals	31	29	60	26	24	50	57	53	110

No significant difference between clinic and control groups.

From these tables and the significant differences found, it is justifiable to draw certain conclusions. There is, first of all, a general tendency for all children to see both the frowning parents as cross with the child in the pictures. Forty-four per cent see the father as cross and 41 % the mother. It is perhaps the case that though a child sees his parents cross about other things, it is only when the anger is vented on the child that it 'registers'. A small group (8·3 % for the father and 10 % for the mother) allow that the father and mother might be cross with each other. According to Dr Kaldegg (4) there is a difference between British and German children in the extent to which they admit parental quarrelling in projection stories, the German children failing almost entirely to accept this. With the present series of pictures parental quarrelling is mentioned by this small group, but the majority accept more easily anger vented on the child. There is no significant difference between groups in this, the children of the control group showing the same general tendency as the clinic group.

The picture of the mother frowning yields a significant difference in that many more of the clinic group see her as sad or miserable than do children in the control group. For instance, a response such as 'she's in trouble—all at once, the little girl was poorly or she died' represents the phantasy of a disturbed child, and this kind of reply is rather more common in the clinical group.

The picture of the child crying throws up several important differences. There is first of all a sex difference in that many more boys than girls see the child crying because punished or scolded by the father. This is not surprising in view of British social life, in which it is quite usual for the punishment of the boy to be left to the father, though the mother may punish the girl herself. More girls see the child as punished by the mother. This sex difference, though present in both groups, is more marked in the clinic group; possibly there is less rigidity in the pattern of discipline in the homes of control group children. In general, boys of the clinical group give significantly more responses that imply punishment by one or both parents or feelings of 'naughtiness' in the child.

Table 4*a*, *b* show a difference between the groups in the attitude to fear. The clinic children see most frequently one or both parents as the source of the child's fright. For the children of the control group, fear is associated much less with parental anger than with some external source, most frequently the fear of animals, especially wolves and dogs, bears and snakes. This difference is not unexpected in view of the fact that in Child Guidance work it is the tensions within the family that generally need easing before much improvement is made, but the verification of the intuitive knowledge is, however, valuable. The difference seems to be not the actual punishment, for both groups project the punishing parent on to the pictures showing frowning adults. But the degree of emotional disturbance the punishment causes seems to be thrown up by the clinic children's attitude to the frightened child. This again is in line with clinical experience, for the well-adjusted child can take even fairly severe discipline and scolding as 'all in the day's work', whereas the maladjusted child tends to be made anxious by comparatively minor punishments.

There is also a sex difference in the fear of animals, the girls being more afraid than the boys. This difference was noted in the writer's earlier work (1), but in neither case does the difference reach a satisfactory level of statistical significance. In the present case P is less than 0·10 but greater than 0·05. The contention, frequently made, that children are made anxious by the cinema or Dick Barton serials on the wireless gains no confirmation here. Only one child mentioned the wireless and one the cinema as a source of fear. Phantasy plays as much round familiar figures of parents and animals as about the ready-made horrors of Grimm or Walt Disney, whose terrors may be real enough while they last, but appear to lack emotional significance over a period. How far the fear of animals is the projection of fear of parental figures into animal forms which can be feared or hated with less anxiety it is impossible to say. It is a little surprising that witches (3), sometimes thought to be widely fear-provoking, are not mentioned at all.

In the final opinion given of the child in the pictures, there is once more a sex difference. Boys tend to classify the child as naughty or bad and in general show more disapproval than girls. Perhaps this links up with the fact, mentioned earlier, that boys see much more the results of anger and punishment in the picture of the child crying and more fear of parents in the picture of the frightened child. It links up also with the writer's finding (1) that in school observations of 'normal' children, boys show more emotional problems than do girls. This does not mean that boys are inherently more unstable than girls, but it may be that present social life makes more demands on boys and creates more tensions in boys than it does in girls. In his study of German prisoners of war, Dr Dicks (2) stresses the unhappy effects on boys of the traditional German ideal of masculinity, with its consequent repression of feminine feeling. Though he has not studied German women in the same way, one may conjecture that the girl, though inferior in status in society, has less of an artificial standard to live up to. To some extent, perhaps, this may be true of our own social conditions.

In the tendency to identify consciously with the child in the pictures, there is no clear-cut difference between groups. About half the children in each group say they are like the child they have described, but some qualify this by saying 'sometimes I'm like that'. Half refuse to identify themselves. The sex difference is not significant.

This test has not been used to any great extent with children in Children's Homes, who have experienced lengthy separation from their parents. In one or two such cases, the

responses have suggested that fear in the test is associated with possible separation rather than punishment, and in a few cases the responses seemed to involve phantasy pictures of rather idealized parent figures. Much more could be done with the responses of larger numbers of children deprived of parental care.

V. Summary and conclusions

1. A series of pictures of the family was designed and given to sixty children attending a Child Guidance Clinic and to fifty 'normal' children in County Schools. These pictures showed a child of the same sex and then the parents first with neutral, then with frowning expressions, the child crying and frightened, a brother or sister and a baby. The child was asked to say what he thought about each picture, no suggestions being made.

2. The tests were analysed in terms of frequency of common responses; the most common response to parental frowning was that the parent was cross with the child. There was no significant difference between the two groups in this respect, nor between girls and boys.

3. A sex difference in the clinic group was that more boys than girls saw the picture of the frowning mother as expressing sadness. This was not true of the corresponding picture of the father.

4. For the picture of the child crying, the most common response implied punishment or scolding by the father. But more boys than girls gave this response, and more girls implied punishment by the mother. Taking together all responses involving parental anger or the child's feeling of 'naughtiness', the boys of the clinic group showed higher frequency of response. The clinic group as a whole showed more punishment responses, but the difference was not statistically significant.

5. For the picture of the frightened child there was a significant difference between the clinic and control groups; the clinic children gave many more responses involving fear of one or both parents' anger or punishment. Children of the control group gave more responses that mentioned fear of animals.

6. There was a sex difference in that boys more often than girls mention fear of the father in accounting for the picture of the frightened child.

7. Boys tended to disapprove more than girls of the child they had described, but there was no significant difference in the extent to which girls and boys would identify themselves with the child in the pictures.

The author would like to thank the teachers who allowed her facilities in schools, Mrs Davies who drew the pictures, and Dr W. H. Whiles for help and advice.

References

(1) Cummings, J. D. (1944). The incidence of emotional symptoms in school children. *Brit. J. Educ. Psychol.* xiv, 156.

(2) Dicks, H. V. (1951). *Psychological Factors of Peace and War.* Ed. Pear. London: Hutchinson and Co.

(3) Hawkey, M. L. (1947). The witch and the Bogey. *Brit. J. Med. Psychol.* xxi, 12.

(4) Kaldegg, A. (1948). Responses of German and English secondary school boys to a projection test. *Brit. J. Psychol.* xxxix, pt. 1, 30.

(5) Raven, J. C. (1950). Comparative assessment of personality. *Brit. J. Psychol.* xl, pt. 3, 115.

(*Manuscript received* 20 *December* 1950)

RESULTS OF A TEST OF HIGH-GRADE INTELLIGENCE APPLIED TO A UNIVERSITY POPULATION

By JEAN G. WALLACE

The Psychological Laboratory, University of Cambridge

I. *Introduction* (pp. 61–62). II. *Results* (pp. 62–68): (i) *the general distribution of scores* (pp. 62–64); (ii) *differences in methods of giving examples* (p. 65); (iii) *correlations between tests and examination results* (p. 66); (iv) *differences between Faculties* (pp. 66–67); (v) *differences between scores of men and women* (pp. 67–68); (vi) *differences between universities* (p. 68). III. *Discussion* (p. 68). IV. *Summary and conclusions* (p. 68).

I. Introduction

In the conclusion to a paper by Dr A. W. Heim, published in this *Journal* in January 1947, and entitled 'An attempt to test high-grade intelligence', some further lines of research were suggested. Some results of this further research are here described.

As stated in Dr Heim's paper, test AH 5 was designed to discriminate between highly intelligent people on grounds of their intelligence. It had been found that with Test AH 4, designed as an intelligence test for use with the general population, the scores of university students fell within the top 30% of those of the general population. Moreover, there were reasons for supposing that a considerable part of any difference shown between the scores of these subjects on such a test was due to differences in speed rather than intelligence.

AH 4 and AH 5, with and without time-limits, were therefore given to 180 Cambridge University students. Results of this testing suggested, among other things, that AH 5 was, in fact, less of a speed test than AH 4, when applied to such a group. It was then proposed:

(*a*) No significant difference having been found between the AH 4 scores of students on the one hand, and lecturers and research workers on the other, that the group of lecturers and research workers should take AH 5, with a view to discovering if their AH 5 scores were significantly higher than those of the students, and thus if discrimination between highly intelligent subjects is demonstrable by means of psychological tests, even where one group is considerably older.

(*b*) That representative groups of students from different universities should be given AH 5, in order to obtain further evidence on any differences between University Faculties. In addition, it was hoped to discover from further student testing any differences in score on grounds of sex or university.

Accordingly, over 1000 students, all volunteers, were tested in 1946 and 1947, the results for 726 on AH 4, 946 on AH 5 and 427 on both tests forming the basis of this paper. These students were of all Faculties. They were taken from Aberdeen, Bristol, Cambridge, Edinburgh, Glasgow, Oxford and Reading Universities, Bedford College, London, Aberdeen Training College and Jordanhill Training College, Glasgow. Results from the two training colleges were included, as the students concerned were all university graduates. Students doing non-degree courses such as Social Science were not included.

In addition, 127 lecturers and research workers took both tests and 44 took AH 5 only. 'Lecturers and research workers' were intended to be a group as nearly as possible *known* to be of high intelligence and assumed, on the whole, to be of higher intelligence

than the average student. The group included professors and lecturers at universities or training colleges for graduates, research workers who had done at least 1 year's research, and those lecturing to or teaching students, though not actually called lecturers —e.g. Teaching Doctors in hospitals. The average age of the group was of necessity somewhat higher than that of the students, and they were divided into two groups for examination of the results—that is, those under and those over 35—as it was thought that differences of score might arise between the older and younger members of this group.

It is not suggested that to be of this standard of intelligence one must be a research worker or lecturer, or even a university graduate, but the criterion used was a fairly easy one on which to assemble a group, and it was felt that such a group was more likely to be comparable with a group of university students on grounds other than intelligence.

Part I of the two tests is verbal and arithmetical and Part II diagrammatic. Both AH4 and AH5 are pencil and paper tests.

II. Results

(i) *The general distribution of scores*

With 726 students Test AH4 gave a range of score of 52 to 127 and a mean of 96·36. The maximum number of marks obtainable is 130, but as the test was designed for use with unselected groups it is not surprising that the scores of such a highly selected group should be bunched at the top end of the range. On AH5, however, designed for use with just such a group, the range of score given by 946 students was 13 to 62 out of a possible maximum of 72 and the mean was 39·06. The distribution curve of the AH5 scores was almost symmetrical. (Figs. 1 and 2 show the distribution curves on the two tests.)

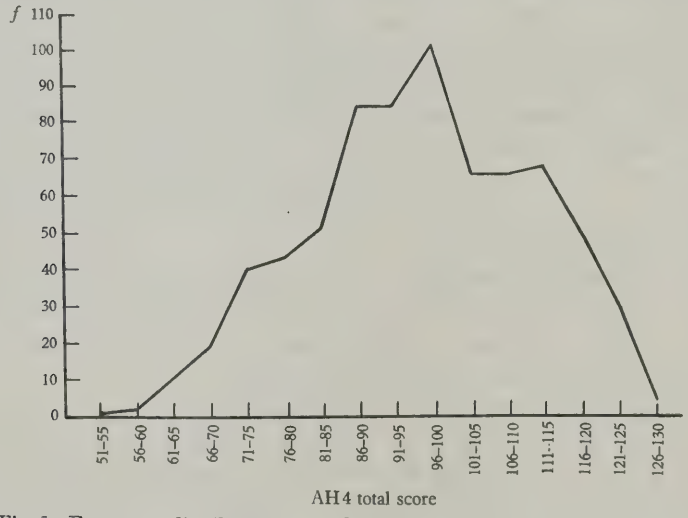


Fig. 1. Frequency distribution curve for 726 students on AH4 total score.

Comparison can be made with a group of 217 naval ratings, whose norms on AH4 were higher than those for the industrial population but by no means so high as those for university students. Though not unselected, they are not highly selected on grounds of intelligence. Their range on AH5 was 6 to 49, with a mean of 23·66, and their distribution curve was skewed towards the lower end. Thus, there seems to be considerable

indication that AH5 is a suitable test for use with a group of high-grade intelligence, and more so than AH4, though not suitable for use with an unselected group.

Scores for 98 lecturers and research workers under 35 years of age ranged from 68 to 129 on AH4, with a mean of 103·77. Those for 29 aged 35 and over gave a range of 53 to 125 and a mean of 94·41. On AH5 the range of score of 123 lecturers and research workers

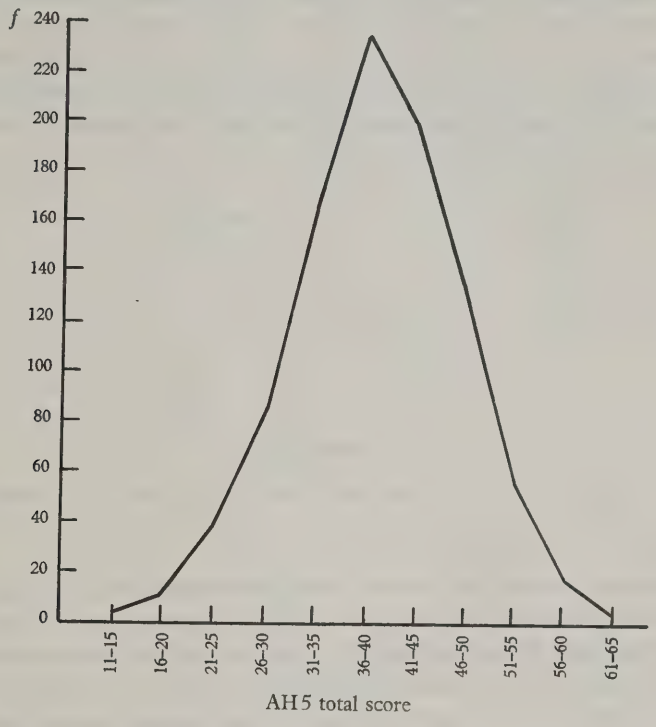


Fig. 2. Frequency distribution curve for 946 students on AH5 total score.

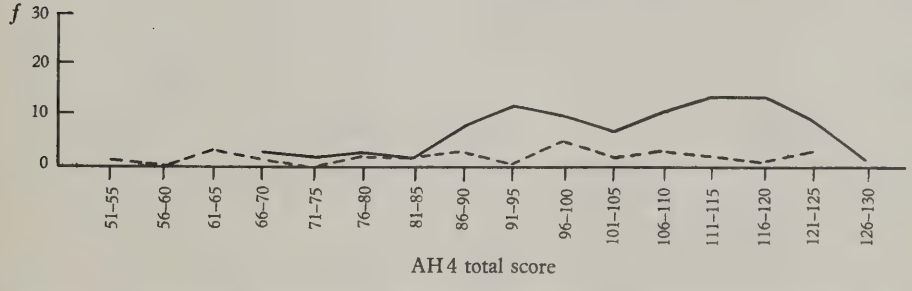


Fig. 3. Frequency distribution curve for lecturers and research workers under 35 (—) and 35 and over (----).

under 35 was 22 to 63 and their mean was 43·18. Scores of 48 aged 35 and over gave a range of 18 to 57 and a mean of 39·87. Thus the scores of the younger group were even more bunched at the top end of AH4 than those of the students, and their distribution curve on AH5 was slightly skewed towards the top end. (Figs. 3 and 4 show the distribution curves for these groups.) It was found, on the whole, that scores on the two tests dropped with increasing age. It is suggested that this is largely a result of the speed factor.

Thus, it may be noted that though the older lecturers and research workers obtain a lower score on AH4 than the students, they do not do so on AH5.

When the scores gained by the students were compared with those gained by the younger group of lecturers and research workers the latter's mean score was not only higher on both tests but the difference between the groups was statistically significant on both tests. Thus, though AH5 does appear to differentiate between the two groups on grounds of intelligence, so, apparently, does AH4, though it is not clear how much the discrimination is on intelligence and how much on speed. It is hoped that further evidence on this point will be produced by an item analysis. However, the curves on the two tests with the group of lecturers and research workers show fairly clearly that AH5 is a more suitable test for such a group.

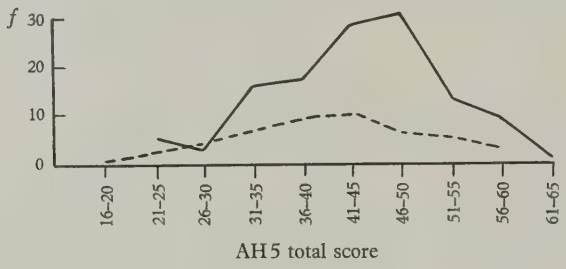


Fig. 4. Frequency distribution curve for lecturers and research workers under 35 (—) and 35 and over (----).

With the students, total scores on the two tests gave a correlation of $r = 0{\cdot}663$, and with the lecturers and research workers the correlation was 0·529. The agreement between the two tests was greatest on Part II. Marks on both tests tend to be higher on this part and it was thought, therefore, that the two parts might be contributing unequally to the total score. This would be so if the standard deviation on Part II was higher than that on Part I. Table 1 shows that this was so in all cases but one (where there were comparatively few subjects) and the difference, though small, is sufficiently consistent to show a weighting in favour of Part II.[1]

Table 1. *Standard deviations on the two parts of the tests*

			Difference	
	Part I	Part II	Pt. II>I	Pt. I>II
Test AH5				
Students				
427 who did AH4 first	4·23	4·95	0·72	—
508 who did AH5 only	4·68	4·94	0·26	—
Lecturers and research workers				
Under 35:				
97 who did AH4 first	4·84	4·89	0·05	—
25 who did AH5 only	5·71	3·42	—	2·29
35 and over:				
29 who did AH4 first	4·48	5·64	1·16	—
19 who did AH5 only	5·35	5·78	0·43	—
Test AH4				
Students				
726 students	2·62	3·18	0·56	—
Lecturers and research workers				
98 under 35	2·55	2·86	0·31	—

[1] I am indebted to Miss V. R. Cane for drawing my attention to the method for deciding this question.

(ii) *Differences in methods of giving examples*

Before each part of each test, examples are provided in order to give the subject some idea of the type of question which he will meet in the test proper, and also to act as a shock absorber for the more nervous subjects. There is no time limit imposed for these examples, and the subjects work them out for themselves, while the tester makes sure not only that all the subjects produce the correct answers but that they appear to understand the reasons for them.

With some of the groups tested in the course of this inquiry, the numbers were too great and the time available was too short for the examples to AH5 to be given in this way. With some of the smaller groups, in such cases, it was possible to allow the students to start on the examples and then to tell them the answers with relevant explanations after 5 or 10 min., by which time the quicker ones had completed them. With the larger groups affected it was not possible to do even this, and the tester had to go through the examples with the group, explaining as well as possible the reasoning behind the answers, but having neither the means nor the time to ensure that the subjects understood the explanations.

Those who had had the answers to the examples entirely dictated to them had a significantly lower mean score on the test than the others, and their scores were therefore excluded from any of the results of the testing. There was no statistically significant difference between the scores of those who did the examples in the usual way and those who had them 'half-dictated', although the latter were lower, so that these scores were included.

It would appear then that the examples play a very important part in the testing. They would seem to be an integral part of the test. The fact that this was shown so clearly when the examples were still included but not worked through by the students themselves is of particular interest in relation to the question of training for intelligence tests. The effect of excluding the examples entirely should be tried.

(iii) *Correlations between tests and examination results*

It was found difficult to obtain examination results which in any way gave an over-all picture of the relative success of a candidate. Where examination results were obtainable they were usually in one subject only or in such a variety of subjects and expressed in such diverse ways that it was almost impossible to compare candidates.

However, examination marks in Psychology throughout the year were available for 29 Edinburgh students, and the mean of these results was compared with AH5 scores, giving correlations with AH5 Part I, Part II and Total score of $r = 0{\cdot}303$, $0{\cdot}230$ and $0{\cdot}312$ respectively. With results on the National Institute of Industrial Psychology's Group Test 33 these examination marks gave a correlation of 0·20. Similar comparisons between AH5 Part I, Part II and Total and Psychology marks for 144 Aberdeen students gave correlations of $r = 0{\cdot}285$, $0{\cdot}176$ and $0{\cdot}266$ respectively. Jordanhill Training College psychology marks for 54 students and AH5 Total score correlated to the extent of $r = 0{\cdot}399$, whereas the same examination results and Moray House Adult Test I scores for 30 of these students correlated to the extent of $r = 0{\cdot}52$. Though low on the whole, all these correlations were statistically significant.

(iv) *Differences between Faculties*

The term 'Faculties' as far as this paper is concerned, has been used to cover four main groups of subjects:

(*a*) Arts—including English, Modern Languages, Classics, History, Law, and degrees such as the Scottish Ordinary M.A. which includes a varied combination of subjects.

(*b*) Sociology—including Economics, Geography, Moral Sciences, Psychology, Commerce and Anthropology.

(*c*) Science and Medicine—including Physics, Chemistry, Agriculture, Pharmacy, Physiology, Zoology, Botany and Biochemistry.

(*d*) Mathematics and Engineering—including Natural Philosophy, Astronomy and Mechanical Science.

It was realized that classification of this sort and involving so many different universities would include some misplacings and would not be entirely accurate, but it was felt that a general division could be made into these four groups.

On AH5 the Part II scores were higher than the Part I scores for the Mathematics and Engineering, Science and Medicine, and Sociology groups, in that order, though none of the differences was statistically significant. The Arts group scored more highly on Part I. On total score there was scarcely any difference between the scores of the groups except for that of the Mathematics and Engineering group who had a significantly higher score than any of the other three groups. This significant difference applied to both parts of the test, except for the score on Part I of the Arts group, which, though still lower than that of the Mathematics and Engineering group, was not significantly so. In the earlier inquiry the Mathematics group was found to have the highest score of all the groups.

When the four main groups were split up into smaller ones according to the actual subjects being studied, the numbers in each group varied from 4 to 285, so that comparison may well be misleading. In addition, these smaller groups were probably even less representative of those doing each subject than the Faculty groups.

However, even when the above points are taken into account, the differences which arise are worth noting. When the mean total AH5 scores were ranked in order for the individual Faculty groups those doing Mathematics were top with a mean score of 45·55, and those doing Agriculture were bottom with a mean score of 36·65. Between these two groups were eighteen others, those doing Physics and Engineering having a mean score of over 41, and those doing Sociology having a mean score of 36·75, and the rest falling between 37 and 39.

Even though methods of selection and numbers probably affect the order of the subjects it is interesting to note that Mathematics and English (which came eighteenth) are in much the same positions as they were in the inquiry on Cambridge students, when the Mathematicians were at the top of the range and those doing English were at the bottom.

Table 2 shows the relative position of the individual Faculty groups on the two parts of AH5. There were two groups which could not be divided further because all that was known of their degrees was that they were 'Arts' or 'Science'. The 'Arts' groups included all those taking an Ordinary M.A. degree at a Scottish University, which involves a fairly wide choice of subjects. These two groups are very large as well as varied, and they have been separated from the rest for these reasons, but it should be noted that

(probably for these reasons) the mean score on both parts of both groups is near the middle of the general range.

It will be seen from Table 2 that the Mathematicians not only led the field on total score but they did so on both parts. There is some evidence that those whose work involves the use of words do better on the verbal and arithmetical part (Part I) and that those studying the more practical subjects (e.g. Mechanical Science, Agriculture, Medicine) do better on the diagrammatic part (Part II).

It should be emphasized that many of the differences between one mean and the next are very small, and that it is unlikely in any results of further testing that the same order would be maintained. It is quite possible, however, that in general the findings would be similar, and it would certainly be most interesting to see how far this is so. The range of mean scores of the groups is quite large and the difference between the top and bottom scores is statistically significant. In addition, in so far as it was possible to compare the groups in this inquiry with those in the inquiry on Cambridge students only, it was found that the different groups maintained the same relative positions to quite a large extent.

Table 2. *Mean* AH5 *scores on the two parts for individual faculty groups*

Part I			Part II		
Subject	No. in group	Mean AH5 score	Subject	No. in group	Mean AH5 score
1. Mathematics	18	21·55	1. Mathematics	18	24·55
2. Classics	16	21·50	2. Physics	17	23·88
3. Modern Languages	24	21·21	3. Engineering	23	22·91
4. Physics	17	20·65	4. Mechanical Science	11	22·18
5. Law	9	20·00	5. Mathematics and other subjects	9	21·89
6. Moral Sciences, etc.	23	19·83	6. Geography	40	21·15
7. Mathematics and other subjects	9	19·78	7. Commerce	4	21·00
8. Engineering	23	19·56	8. Moral Sciences, etc.	23	20·13
9. English	39	19·43	9. Agriculture	26	20·00
10. History	29	18·96	10. Medicine	81	19·81
11. Economics	41	18·68	11. Classics	16	19·31
12. Mechanical Science	11	18·64	12. Modern Languages	24	19·00
13. Geography	40	18·20	13. Chemistry	15	18·93
14. Chemistry	15	18·13	14. Economics	41	18·90
15. Sociology, etc.	24	18·12	15. History	29	18·62
16. Commerce	4	18·00	15. Sociology, etc.	24	18·62
17. Medicine	81	17·59	17. Law	9	17·78
18. Agriculture	26	16·65	18. English	39	17·59
Arts, etc.	285	19·51	Science etc.	185	21·13
Science, etc.	185	19·31	Arts, etc.	285	19·36

(v) *Differences between scores of men and women*

On the assumption that methods of selection of both men and women are much the same—if not entirely so—at universities other than Oxford and Cambridge, such universities were taken here as one group and Oxford and Cambridge, where the case is different, were taken as another.

With the students in the first group, on AH4 total score, the 246 men obtained a mean score of 99·30 and the 291 women a mean score of 93·97, this difference being statistically significant. On AH5 total score, the 286 men concerned had a mean score of 40·66 and the 333 women a mean score of 37·59. This difference was also statistically significant.

Thus, on both tests, the men scored significantly more highly than the women, and so far as can be ascertained, this result is not due to differences in selection.

When Oxford and Cambridge—where the selection of women undergraduates is at a highly competitive level—were taken as another group, the scores on the two tests showed differences in the expected direction. AH5 scores only were used, as none of the Oxford students had taken AH4. The men obtained a mean total score on AH5 of 39·56, and the women a score of 38·00. This difference was not significant. The numbers in this group were much more uneven, with 269 men and 53 women.

The greater part of the differences in total score with the first group was accounted for by the higher score obtained by the men on Part II.

(vi) *Differences between universities*

No significant differences were found between the mean scores of the different university groups, a fact which is of interest even though these groups are not necessarily representative and were in some cases drawn from only a few Faculties.

III. Discussion

From the distribution of scores with subjects of high-grade intelligence (in particular in comparison with AH4 scores), AH5 appears to be a suitable test for such a population. This was confirmed by the fact that the subjects came from so many different universities and by the lack of any difference in score between the university groups. It is only when the results are studied in more detail that some doubts arise as to whether the construction of the test may be affecting them. In particular, it has been shown that the Part II scores contribute more to the total score than the Part I scores, and since students studying the more practical and non-verbal subjects gain higher marks on Part II, the test would seem to be biased in their favour, the same being the case with the men as opposed to the women.

The other main point arising from these results is the effect which the small amount of experience given by the examples has on test scores—of particular importance to the question of training for and unequal experience in the taking of intelligence tests.

IV. Summary and conclusions

1. Over 1000 students, volunteers drawn from eight universities and two teachers' training colleges, and 171 lecturers and research workers were given either or both of two intelligence tests—AH4 and AH5—primarily in order to gain further evidence on the value of a test designed for use with subjects of high-grade intelligence.

2. Test AH5 (the test of high-grade intelligence) gave a much better distribution curve than did Test AH4 (an intelligence test designed for use with an unselected group). The curve showed that the scores of the students and lecturers and research workers were spread evenly over Test AH5, whereas on AH4 they were limited to the top end of the possible range.

3. The mean score gained by the group of lecturers and research workers was significantly higher than that of the students on both tests. Assuming that the group of lecturers and research workers were of higher intelligence than the group of students, both

tests would appear to differentiate between the groups on grounds of intelligence, unless both (or one) still differentiate between them on grounds of speed.

4. The greatest agreement between the tests was on the diagrammatic part. Scores on the whole are higher on this part and it appears to contribute more to the total score than the other.

5. Those students who had the answers to the examples to AH5 dictated to them, with relevant explanations, instead of having time to work them out for themselves, obtained a significantly lower score on the test than those who were able to do them in their own time. (Their test scores were therefore excluded from the results.)

6. Correlations between AH5 scores and the very limited examination results available were significant but low.

7. Those students studying Mathematics and Engineering scored significantly higher marks on AH5 than students in any other Faculty groups, except on Part I, where there was no significant difference between this group and those studying Arts subjects.

8. When the Faculty groups were further subdivided into actual subjects studied, there was a difference of ten marks between the top and bottom groups, and some evidence that those whose work involves the manipulation of words do better on Part I and that those whose work is more practical do better on Part II.

9. In the group of universities other than Oxford and Cambridge, men were found to have significantly higher scores on both tests than women. Though the scores of the men from Oxford and Cambridge were higher than those of the women, the difference was not statistically significant. The higher scores gained by the men were due largely to the higher scores gained by them on Part II.

10. There was no significant difference between the scores of students from the different universities.

11. It is suggested that some of the differences in score may be due to the design and construction of the tests.

I am indebted to Dr A. W. Heim for initiating and giving help in this research and for the use of the tests, and to Prof. Sir Frederic Bartlett for help and guidance throughout. In addition, I would like to thank all those who gave facilities for the testing to be done, the students, lecturers and research workers who volunteered as subjects, and those who have so much help in arranging the testing and in obtaining examination results and other data.

(*Manuscript received* 23 *May* 1951)

MANUAL TRACKING ON A HORIZONTAL SCALE AND IN THE FOUR QUADRANTS OF A CIRCULAR SCALE

By NORAH E. GRAHAM

The Nuffield Department of Industrial Health, University of Durham
(King's College, Newcastle upon Tyne)

I. *The problem* (p. 70). II. *Exp.* 1 (pp. 71–73): (i) *Method* (p. 71); (ii) *Results* (p. 71). III. *Exp.* 2 (pp. 73–75): (i) *Method* (p. 73); (ii) *Results* (p. 73). IV. *Discussion* (pp. 75–77). V. *Summary* (p. 77). *References* (p. 77).

I. The problem

A preliminary tracking experiment has already been described (1), the purpose of which was to study the effect of the shape of a display on the subject's ability to keep a moving pointer centred on a given scale reading. Three targets were chosen for comparison, one at the mid-point of a horizontal scale, one at the mid-point of a vertical scale and another in the lower segment of a circular scale. A circular control knob was used throughout, and it was found that the tracking error was significantly smaller on the horizontal scale than on either of the other two displays.

The question which now arises is this: How far does the choice of the target within the main display influence the results? Presumably, from the subject's point of view, the problem would be exactly the same at any other comparable target on the linear scales as the relation between control and display is constant. This is not so, however, on the circular scale. In a tracking task the subject concentrates on the particular sector of the circle in which the target lies. Thus, as the control is turned in a clockwise direction, the pointer travels from left to right in the upper quadrant, down the right-hand side, from right to left in the lower quadrant and up the left-hand side (Fig. 1). Bearing this in mind it may be that the previous results relating to the circular scale were due to the choice of target.

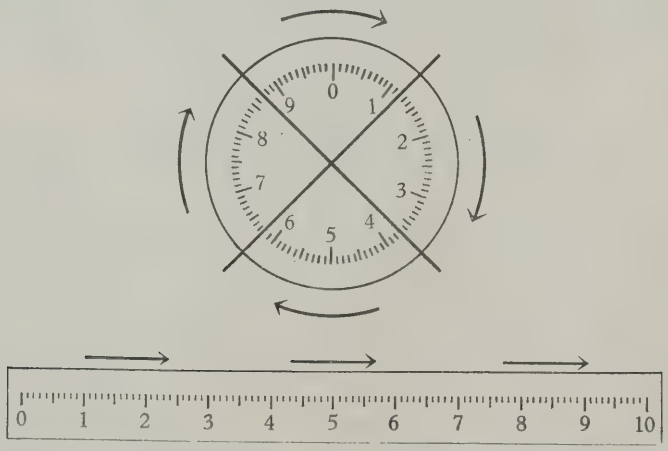


Fig. 1. Diagram to show the horizontal scale and the circular scale which is divided into the four quadrants. The arrows indicate the direction of pointer movement when the control knob is turned clockwise.

Two further tracking experiments were therefore planned. The first was designed to compare the speed and accuracy of tracking on the horizontal scale with two diametrically opposite targets, one in the upper and one in the lower quadrant of the circular scale. In the second experiment one target was chosen in each of the four quadrants of the circular scale, in order to find out how these changes in display-control relation affect the subject's performance.

II. Exp. 1. To compare the speed and accuracy of manual tracking on a horizontal scale with targets in the upper and lower quadrants of a circular scale

(i) *Method*

The apparatus used for the tracking experiments has been described fully in a previous paper (1). The subject sits in front of the display and by means of a circular control knob, situated vertically below the display, attempts to counteract the intermittent disturbances produced by a cam drum.

Eighteen male students were the subjects, and the problem was described to them in groups of six. To avoid bias no mention was made of the results of the previous experiment. They were tested individually, and each man in a group tackled the three tests in a different order. The targets chosen were '5' on the horizontal scale and '4' and '9' on the circular scale. In what follows these will be denoted by the symbols H_5, C_4 and C_9 respectively.

Before each test the subject was allowed to handle the control knob for 15 sec. to learn the relation between direction of rotation of control and pointer movement on the display. This was followed by 45 sec. practice, then the test run which lasted $1\frac{1}{2}$ min. In between tests the subjects had a few minutes' rest while the display was changed.

The method of scoring which has previously been described was adopted, i.e. the tracking error was measured in division-seconds and a record was kept of the number of times the control was turned the wrong way. The first and last quarter minutes of the test run were not included in the score.

(ii) *Results*

The basic data are shown in Table 1.

As in the first tracking experiment the individual tracking error bears no simple relation to the number of times the control is turned the wrong way. Some of these subjects, notably nos. 15, 17 and 18, maintained consistently low error scores in spite of repeatedly turning the control the wrong way. In these cases the initial mistake is corrected at once and followed by an abnormally quick response. Subject no. 1, on the other hand, took a long time to respond to each disturbance, as indicated by the high tracking error, but the response, when it came, was always the right one.

The mean scores and their standard errors are given in Table 2, and it will be seen that in spite of these individual characteristics the target which gives the highest tracking error is also the one at which most mistakes are made in the use of the control.

These results suggest that there is very little difference between a target on the horizontal scale and C_9 in the upper segment of the circular scale. This is to be expected as the control-display relation is comparable. At C_4, however, where the sense relation is reversed, the tracking error is greater, and the control is more frequently turned the wrong way. As in the previous experiment, there is only a slight tendency to improve after the first test.

Table 1. *To show, for each subject, the order of presentation of the displays, the tracking error and the frequency of turning the control the wrong way*

	Order of presentation			Tracking error (divisions × seconds)			Control turned the wrong way (no. of times)		
Subject	H_5	C_9	C_4	H_5	C_9	C_4	H_5	C_9	C_4
1	1	2	3	220·2	229·5	282·8	0	0	0
2	2	3	1	168·0	173·4	182·0	0	2	2
3	3	1	2	117·5	125·3	133·9	1	3	1
4	3	2	1	181·6	191·8	196·7	3	2	1
5	1	3	2	127·7	151·5	136·6	5	3	1
6	2	1	3	153·5	148·8	128·6	4	1	2
7	1	2	3	171·8	136·2	201·2	10	12	13
8	2	3	1	150·8	136·4	148·7	1	1	3
9	3	1	2	126·7	185·0	164·0	2	1	6
10	1	3	2	160·4	122·2	154·0	1	2	4
11	2	1	3	146·2	164·4	152·6	3	5	4
12	3	2	1	107·8	106·3	152·9	5	10	13
13	1	3	2	138·8	144·1	134·4	0	3	2
14	2	1	3	142·0	150·6	156·0	5	0	1
15	3	2	1	83·2	93·2	120·7	4	4	5
16	1	2	3	157·9	93·8	135·4	5	2	8
17	2	3	1	116·7	118·2	129·7	7	9	9
18	3	1	2	110·2	112·2	126·5	4	2	3

H_5 = horizontal scale, target 5; C_9 = circular scale, target 9, etc.

Table 2. *Mean scores for eighteen subjects*

	Tracking error (divisions × seconds)			Control turned the wrong way (no. of times)		
Target	Mean	±	S.E.	Mean	±	S.E.
H_5	143·4		7·5	3·3		0·63
C_9	143·9		8·5	3·4		0·79
C_4	157·6		9·2	4·3		0·94
Order of presentation						
1st	155·2		6·7	3·7		0·88
2nd	143·2		7·6	3·7		0·76
3rd	146·1		10·5	3·7		0·78

Table 3. *Analysis of variance of the data given in Table* 1

Source	Sums of squares	Degrees of freedom	Variance	F	P
Tracking error:					
Between persons	54,501	17	3206	11·2	$\leqslant 0{\cdot}001$
Between targets	2,404	2	1202	4·2	$\leqslant 0{\cdot}05$
Between order	1,418	2	709	2·5	Not significant
Residue	9,122	32	285	—	—
Total	67,445	53	1273	—	—
Control turned the wrong way:					
Between persons	484·6	17	28·5	8·6	$\leqslant 0{\cdot}001$
Between targets	10·8	2	5·4	1·64	Not significant
Residue	111·9	34	3·29	—	—
Total	607·3	53	11·46	—	—

The variations due to the individual differences, the change of target and the order in which the tests were carried out, are analysed in Table 3.

As before, the greatest source of variation is the individual ($P \leqslant 0{\cdot}001$). However, the differences in the tracking errors due to the position of the targets are significant at a 5% level of confidence. It can be shown, by an application of the t test, that the difference between the means must exceed 11·9 div.-sec. in order to show this significance

The conclusions reached in this experiment therefore confirm those previously put forward, namely, that the tracking error is significantly smaller on the horizontal scale than at a target in the lower quadrant of the circular scale. It is now suggested, however, that the position of the target on the circular scale is the determining factor. For the purpose of tracking, the difference between the horizontal scale and the upper quadrant of the circular scale is negligible compared with the difference between two diametrically opposite targets on the circular scale.

In a second experiment four targets on the circular scale were compared.

III. Exp. 2. To compare the speed and accuracy of manual tracking in the four quadrants of the circular scale

(i) *Method*

The four targets chosen on the circular scale were 0, 2·5, 5 and 7·5 (Fig. 1). As the scale markings representing the whole numbers are slightly longer than those representing the subdivisions, it was thought that 0 and 5 might form easier targets than 2·5 and 7·5. In order to make the four positions comparable, therefore, the dial markings at 2·5 and 7·5 were increased slightly in length until they were equal to those at 0 and 5. Twenty-four subjects took part in the experiment, all post-graduate engineering students. Each subject was presented with the four targets in a different order in an effort to eliminate the effects of learning or fatigue. The test routine was exactly as described in the previous experiment.

(ii) *Results*

Table 4 contains the basic data, and the means and their standard errors are shown in Table 5.

Table. 4. *To show, for each subject, the order of presentation of the targets, the tracking error and the frequency of turning the control the wrong way*

	Order of presentation				Tracking error (division × seconds)				Control turned the wrong way (no. of times)			
Subject	C_0	$C_{2·5}$	C_5	$C_{7·5}$	C_0	$C_{2·5}$	C_5	$C_{7·5}$	C_0	$C_{2·5}$	C_5	$C_{7·5}$
1	1	2	3	4	169·0	127·9	125·4	130·6	0	4	1	0
2	1	2	4	3	137·4	138·3	149·6	142·0	3	7	8	5
3	1	4	3	2	207·3	225·3	222·5	194·9	0	0	0	0
4	1	3	4	2	142·8	140·3	171·6	150·4	0	0	0	0
5	1	3	2	4	143·7	182·3	233·0	170·3	3	4	4	4
6	1	4	2	3	111·8	113·7	148·0	132·5	3	4	3	4
7	2	1	3	4	103·7	124·2	122·4	133·4	1	3	8	10
8	2	1	4	3	149·6	179·4	141·3	125·1	2	4	1	2
9	3	1	2	4	149·6	186·9	183·9	256·5	3	2	4	5
10	4	1	2	3	180·0	177·1	245·1	216·7	1	2	0	1
11	3	1	4	2	134·8	168·3	144·0	147·8	2	8	7	5
12	4	1	3	2	184·9	246·9	173·5	191·2	4	2	1	4
13	2	3	1	4	112·7	103·5	117·6	134·7	3	3	2	4
14	2	4	1	3	260·8	312·9	262·8	293·2	0	2	0	4
15	3	2	1	4	137·5	156·3	228·1	170·2	3	0	3	0
16	4	2	1	3	114·7	129·5	164·5	201·3	4	4	8	3
17	3	4	1	2	265·9	205·4	237·7	228·8	0	0	0	1
18	4	3	1	2	112·1	130·5	154·2	129·5	5	6	2	8
19	2	3	4	1	105·1	131·7	127·4	131·6	2	0	2	1
20	2	4	3	1	135·1	171·0	193·3	185·8	0	1	4	1
21	3	4	2	1	127·1	101·1	110·3	136·0	0	0	1	0
22	4	3	2	1	147·1	170·1	155·3	176·2	0	5	4	5
23	3	2	4	1	102·8	102·1	123·2	106·5	0	2	1	3
24	4	2	3	1	130·9	207·9	194·8	192·4	4	8	5	6

$C_{2·5}$ = circular scale, target 2·5, etc.

When the pointer is kept on C_0 the tracking error is least and the smallest number of mistakes are made in the use of the control. This is interesting as this target was chosen deliberately for its similarity to a target on the horizontal scale. Again the characteristic improvement in the tracking error after the first test is evident but beyond this there appears to be no further improvement with practice. Table 6 shows an analysis of variance of the results.

Table 5. *Mean scores for twenty-four subjects*

	Tracking error (divisions × seconds)			Control turned the wrong way (no. of times)		
Target	Mean	±	S.E.	Mean	±	S.E.
C_0	148·6		9·04	1·79		0·33
$C_{2.5}$	163·8		10·37	2·96		0·52
C_5	172·1		9·29	2·87		0·55
$C_{7.5}$	169·9		9·39	3·17		0·69
Order of presentation						
1st	170·3		8·8	2·54		0·48
2nd	160·3		9·64	2·79		0·51
3rd	163·3		10·11	2·67		0·46
4th	160·5		10·14	2·79		0·58

Table 6. *Analysis of variance of data given in Table* 4

	Sums of squares	Degrees of Freedom	Variance	*F*	*P*
Tracking error:					
Between persons	164,324	23	7144	13·5	$\leqslant 0{\cdot}001$
Between targets	8,073	3	2691	5·1	$\leqslant 0{\cdot}01$
Between order	1,590	3	530	1·0	Not significant
Residue	34,884	66	528	—	—
Total	208,871	95	—	—	—
Control turned the wrong way:					
Between persons	351	23	15·3	5·5	$\leqslant 0{\cdot}001$
Between targets	27	3	9·1	3·2	$\leqslant 0{\cdot}05$
Residue	192	69	2·8	—	—
Total	570	95	—	—	—

Once again the most striking differences are those which occur between one individual and another ($P \leqslant 0{\cdot}001$). The differences due to the targets, however, are sufficiently great to say that it is unlikely that they are merely chance variations. To be significant at the 1% level the difference between the mean tracking errors for two targets must exceed 18·7 div.-sec. A difference of 13·8 div.-sec. is significant at the 5% level. It may therefore be concluded that the error at C_0 is significantly less than at any of the other three targets. This is also true of the number of times the control is turned the wrong way. Significantly fewer mistakes are made at C_0 than at any other target ($P \leqslant 0{\cdot}05$). All other differences may well be due to chance.

The results of these two experiments are shown graphically in Fig. 2. To complete the picture the results of the original tracking experiment in which horizontal, vertical and circular scales were compared, have also been included. Care must be taken, however, in making comparisons between two different experiments as the subjects were not the same and individual differences in tracking ability are very great. Nevertheless, the superiority of the positions H_5, C_9 and C_0 is very apparent.

Another point which the diagram shows is the positive correlation between the tracking error and the number of times the control is turned the wrong way. It is important at

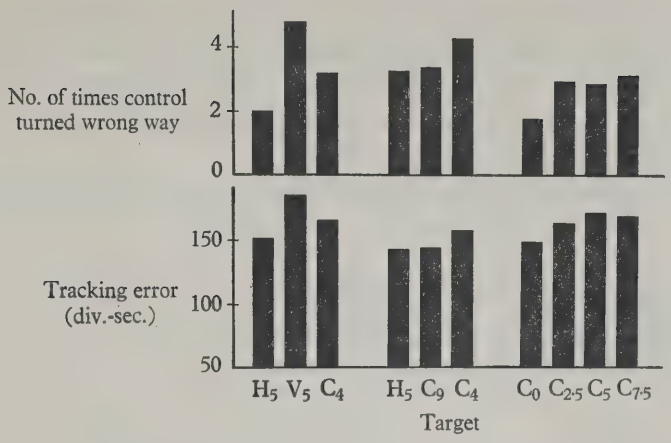


Fig. 2. The results of three tracking experiments. H_5 = horizontal scale, target 5, V_5 = vertical scale, target 5, $C_{2\cdot5}$ = circular scale, target 2·5, etc.

is stage to inquire whether this is bound to arise, from the very nature of the graphic cord, or whether these two criteria can be regarded as independent pieces of evidence. n analysis of covariance shows that when considering the average scores of all the bjects at the different targets there is, as the diagram suggests, a high positive correlation between the two types of score ($r = +0{\cdot}92$). But the *individual* who makes any mistakes in handling the control may, at the same time, have a comparatively nall tracking error. In this case the correlation is not significant ($r = -0{\cdot}26$). The low lue of r ($+0{\cdot}12$) associated with the residual term in the analysis shows that the area der the curve is an independent variable and is not dependent upon the number of nes the control is turned the wrong way. These two aspects, therefore, of the human sponse seem to contribute independent pieces of evidence in favour of the targets $_5$, C_9 and C_0.

IV. Discussion

ny experiments on instrument reading which require a manual response must give equate consideration to the rôle played by this response in influencing the results.' is was a conclusion reached by Grether & Connell (2), and its truth is borne out by e present investigation. The original purpose of these tracking experiments was to k for differences which may exist in the human response to horizontal, vertical and cular scales. Since the method of measuring the response involves the use of a manual ntrol, the interrelation of control and display is important. It may favour the performance on one display at the expense of another and any interpretation of the results ust be in terms of the display-control relation rather than in terms of the display alone. The tests with various targets on the circular scale as well as those on the linear scales point to one conclusion. It seems quite evident that the subject associates clockwise ation of a control knob most readily with pointer movement from left to right, either ng a linear scale or along the arc in the upper segment of a circular scale. This idea shown diagrammatically in Fig. 3. This conclusion only holds, however, when the ntrol is vertically below the display. Whether or not it is true with other arrangements control and display cannot be said.

Warrick (3, 4) has also studied the relation of rotary controls to linear and semicircular displays. In his experiments the subjects used the control to move a light on the display to a central position. Two main conclusions have been drawn from his work, namely:

(*a*) The edge of the control nearest to the display should move in the direction in which the indicator is to be moved.

(*b*) Clockwise rotation of a control should be associated with clockwise movement on an indicator.

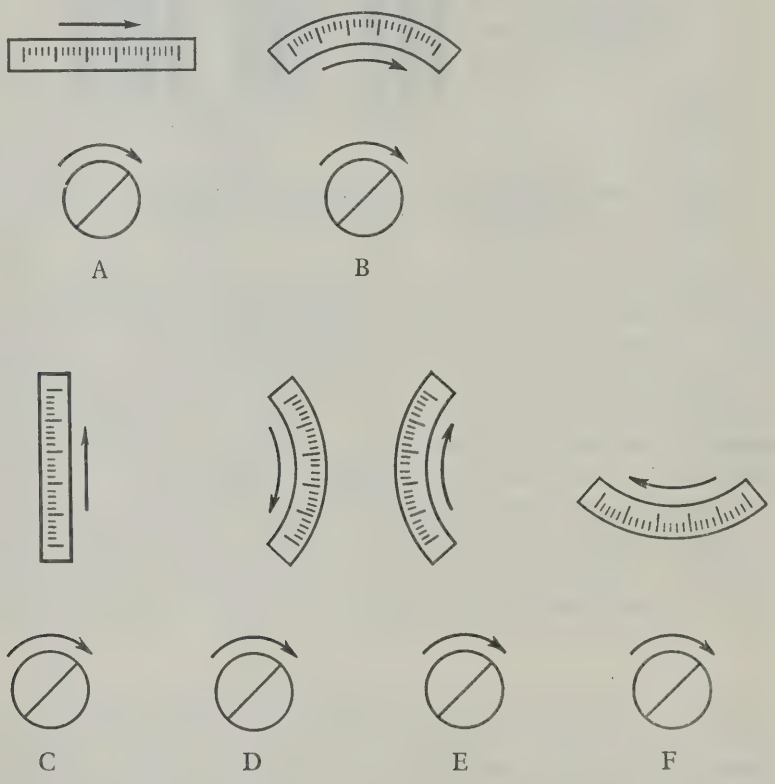


Fig. 3. Relation of pointer movement to control knob movement. Experiments show that A and B are better than C, D, E, or F.

It will be seen from Fig. 3 that the two combinations of display and control that proved best in the present work *fulfil both these conditions*. In the lower quadrant of the circular scale, however, it is only possible to fulfil one condition or the other. As Diagram 'F' in Fig. 3 shows, clockwise rotation of the control does produce clockwise movement on the indicator, but to satisfy the first condition it should have the reverse effect and move the pointer from left to right. Warrick himself found that with such an arrangement of display and control his results were difficult to explain. There is some evidence from his work to suggest that when only the lower segment of a circular scale is used, and the control is situated below the display, it might be better to associate clockwise rotation of the control with pointer movement from left to right on the scale, even though this is anticlockwise. Further experiments are necessary to clarify this

oint which is a display problem of interest to the designers of pressure gauges since ıey, among others, are faced with a situation in which the control must be turned ockwise to decrease the scale reading.

V. SUMMARY

1. A comparison is made of the speed and accuracy of manual tracking on a horizontal ;ale with two targets on a circular scale. One of these is in the upper and one in the ›wer quadrant of the circle.

2. There is no difference between the subject's performance on the horizontal scale ıd at a target in the upper quadrant of the circular scale. This is attributed to the milarity of the display-control relation. In the lower quadrant of the circular scale, here the sense relation is reversed, the tracking error is significantly greater and the ›ntrol is turned the wrong way more often.

3. Four targets are compared, one in each quadrant of the circular scale, to determine ›w the variations in display-control relation affect the speed and accuracy of performance.

4. The tracking error is least and the control is turned the wrong way less frequently the upper quadrant of the circular scale. There is no significant difference between ıe mean scores in the other three quadrants.

5. These results suggest that a person associates clockwise rotation of a control knob ıost readily with pointer movement from left to right either along a linear scale or ong the arc in the upper segment of a circular scale. This, however, has only been ;monstrated when the control is vertically below the display.

Acknowledgements are due to Prof. R. C. Browne for advice in the work and in the :eparation of this paper, and to the Medical Research Council for providing him with grant for scientific assistance; also to Mr H. Campbell, B.A., F.S.S., for statistical help.

REFERENCES

) GRAHAM, N. E., BAXTER, I. G. & BROWNE, R. C. (1951). *Brit. J. Psychol.* (General Section), XLII, 155.
) GRETHER, W. F. & CONNELL, S. C. (1948). U.S. Air Force. Air Material Command. Wright-Patterson Air Force Base. Memorandum Report MCREXD–694–17A.
) WARRICK, M. J. (1947). *Psychological Research on Equipment Design*. Report, No. 19, p. 137. Ed. Fitts P. H. Washington, U.S.: Government Printing Office.
) WARRICK, M. J. (1949). U.S. Air Force. Air Material Command. Wright-Patterson Air Force Base. A.F. Technical Report, No. 5812.

(*Manuscript received* 13 *September* 1951)

PUBLICATIONS RECENTLY RECEIVED

Aspects of Form. Edited by Lancelot Law Whyte. London: Lund Humphries. 1951. Pp. ix+249. 21*s.* net.

This interesting book had its germ in the catalogue of a recent exhibition held in London by the Institute of Contemporary Arts, and dedicated to the theme of 'Growth and Form'. It is a symposium comprising eleven essays contributed by an equal number of writers representing various fields of study ranging from physics through biology and psychology to aesthetics; and the common subject is Form in its manifold aspects. Actually only spatial form is taken into consideration; but this in itself is a vast subject, including as it does both external form (or shape) and internal form (or structure). The aim of the symposium is to direct attention to the universal and central importance of problems concerning Form in every branch of human inquiry. It would be misleading to infer from the title that this volume is but a collection of treatises from a modern Academy of Plato; for, in spite of differences of approach, each contribution is inspired by the spirit of the most recent scientific outlook. An increasing recognition of the significance of pattern is, of course, characteristic of this outlook; but equally characteristic is the recognition of transformation of pattern. The two most specifically psychological essays in the symposium (those by Lorenz and Arnheim) are, almost inevitably, concerned with Gestalt psychology. These, together with Waddington's essay on biological form and that of Gombrich on artistic form, will doubtless prove of greatest interest to the psychological reader. It is claimed by the Editor that this book is probably the first devoted to a general survey of Form in physics, biology, psychology and art; and it is not only interesting on this account, but also challenging to thought on account of the prospect which it opens out for the eventual relation of the various discrete sciences to one another. W.R.D.F.

La Psychologie de l'Art dramatique. By André Villiers. Paris: Librairie Armand Colin. 1951. Pp. 224. 200 fr.

This book is remarkable for the skill with which the author has assembled and balanced his vast knowledge of all aspects of his subject. A brief historical survey quickly establishes his lively independence of judgement. He goes on to consider and interrelate the problems and contributions of the theatre architect, the author, the producer, the scene designer, the stage manager and the audience.

Whilst it is primarily a book for all those interested in the theatre, enthusiasts for psychodrama—it would be interesting to know what other drama there is—may gain some fresh insight into the problems of expressive movement, empathy, sublimation, self-display and even the more general one of perception. It is doubtful whether the author's discussion of these problems would greatly enhance the psychologist's knowledge, but it might quicken his awareness of the fact that a field for the study of affective organization far richer than the cinema has been curiously neglected. G.A.F.

The Art of Scientific Investigation. By W. I. B. Beveridge. London: Heinemann. 1950. Pp. xii+171. 10*s.* 6*d.* net.

This book is intended primarily, though not exclusively, as a guide to the student about to engage in scientific research. The author is a biologist, and he has biological research chiefly in mind; this must be remembered when one meets with passages that might seem questionable from the point of view of the physical research worker. The book is well written, is full of interesting examples of the matters discussed, and covers the chosen field almost completely. Indeed, the only point that seems to be omitted, or almost so, is the function of scientific societies and international organizations in the furtherance of research. This is rather strange, for the author certainly knows the part they play, yet even in the passage on publication he writes as though the research worker alone decides what is to be published. In such a book there is inevitably much that is platitudinous, but Prof. Beveridge writes throughout from experience and that makes even platitudes live. The book is timely, it fills a gap in the literature of research, and can be recommended to all who are intending to take up biological investigation.

This being understood, one or two small criticisms might perhaps be permitted. Prof. Beveridge pays lip service to the study of the history of science, but when he advises against the study of German 'until German science is properly revived', since the German output has been very small 'during the last ten

ears', it needs little psychology to deduce that the classics of the past, on which the validity of present ience rests, have little interest for him. Again, while one may heartily agree that the study of the philo- phy of science is of little help in research—as little, in fact, as the study of the traffic laws is in breaking ie speed record—yet there is something more to be said than he admits in favour of the view that the udent engaged in doing things should have some idea of what he is doing. The chapter on the place of ason in research reflects the author's somewhat excessive indifference to this subject. Much of what he ys about the misapplication of reason in research is undoubtedly true, and he does not fail to point out at nevertheless reason has its proper place, but it is doubtful if he has realized exactly what and how portant that place is. We would guess, for instance, that he would hold that scientific hypotheses are sted by their accordance with the facts, but that is not so; if they did not agree with the facts they ould not have been framed. They are tested by accordance of facts with *logical deductions* from them—so uch so, indeed, that many hold the chief function of hypotheses to be that of leading to such deductions. r Arthur Eddington (speaking, it is true, of physics and not of biology) went so far as to claim that the hole scheme of physical law could on principle be deduced by pure reason without any observation at l. This we believe to be an exaggeration, but it is certainly true that, to an imaginative and logical mind, very little observation can go a very long way by the right use of reason. One might also add a plea for eater recognition of the importance of imagination in the literal sense of the word—i.e. the power of rming vivid images in the mind which was so characteristic of Faraday and Rutherford, for example. cof. Beveridge writes of imagination as though it were indistinguishable from intuition.

All these, however, are minor points. The book as a whole is a most stimulating one and deserves to be idely read. The number of misprints is excessive, but they are harmless. HERBERT DINGLE

rogressive Matrices (1947), *Sets A, Ab, and B. The Crichton Vocabulary Scale.* By J. C. RAVEN. London: H. K. Lewis and Co. Ltd. and George G. Harrap and Co. Ltd. 1951. Specimen Set with Instructions, etc., 17*s.* 6*d.* net.

his Progressive Matrices Scale consists of sections *A* and *B* of the 1938 test with a new *Ab* section of velve problems sandwiched between them. The purpose is to provide a scale giving finer discrimination the lower end for use with children under the age of 11 and for clinical work. The accompanying manual scribes two versions of the test, of which the first is a 'formboard' requiring the actual placing of a piece complete the pattern. It does not appear that this is at present available for purchase. The book rsion is similar to the original scale, but the problems (except for the last few) are printed, and quite well inted, in colours. An improvement is that the problems are printed on one side only of each page, and e paper is of reasonable opacity so that succeeding diagrams do not show through the page presented fficiently to distract. The original instructions in the manual for administering the book form of the st were found to be inadequate, and revised directions which should be clear and precise enough are ven on an enclosed leaflet. This also presents the latest norms—in percentile form—based on some 0 Scottish schoolchildren. The main manual contains an interesting discussion on the evaluation of the st responses.

The Crichton Vocabulary Scale (1950) similarly derives from the definitions form of the Mill Hill ocabulary Scale. It consists of two lists of 40 words each. The first is made up of the first 20 words om each of the Mill Hill *A* and *B* lists with some rearrangement of order: the second is new. Both lists e to be given. Administration is similar to the Terman-Merrill method, and full directions together th samples of acceptable and unacceptable responses for each word are given in the manual. Norms e presented in percentile form. Test-retest reliability coefficients for the Crichton scale are given as 37 at age $6\frac{1}{2}$, 0·95 at $9\frac{1}{2}$ and 0·98 over the age range 5–10, when the test is repeated after 2 months. rrelations with the Progressive Matrices are, respectively, 0·53, 0·65 and 0·94; with Terman-Merrill 34 at age $6\frac{1}{2}$ and 0·79 at $9\frac{1}{2}$.

The Crichton Vocabulary Scale and the present Progressive Matrices Scale are intended for use together afford separate measures of 'present capacity for intellectual activity' and 'general cultural attain- ents'—a proceeding which, as is well known, Raven prefers to a single test of 'general intelligence'. The ichton manual discusses the significance of discrepancies between grades obtained on the two tests. ese should prove useful instruments, and though they are unlikely at present to oust Terman-Merrill m the affections of most psychologists who work with young children, they are valuable alternative ts, particularly for handicapped children and will, no doubt, be most valuable for clinical use and search purposes. A.H.D.T.

Theory of Mental Tests. By H. Gulliksen. London: Chapman and Hall. 1950. Pp. xix + 486. 48*s*.

In this book are gathered together most of the equations and formulas dealing with test scores which have been published during the last fifty years. The derivations and limitations of various measures of test 'reliability' and 'validity' are dealt with extensively and such topics as the effect of test length on reliability and validity, methods of scoring tests, the weighting of scores, item analysis, etc., are fully discussed. Given certain assumptions, which are scrupulously pointed out whenever they are made, the algebra is worked out clearly enough, but how many psychologists would be willing to make these assumptions? The basic assumption is that a person's raw score on a test consisting of a number of different items is composed of a 'true' score and an 'error' component. The 'true' score is unknown but is taken to represent the 'actual ability of the person, a quantity that will be relatively stable from test to test as long as the tests are measuring the same thing'. The 'true' and 'error' scores are assumed to be uncorrelated. On this simple basis is built a complex algebraic edifice, buttressed here and there by further assumptions such as, for example, the existence of 'parallel tests', of which the following statements are made: 'The means and standard deviations of parallel tests are equal', and 'all intercorrelations of parallel tests are equal'. The standard deviation of the unknown 'error' score is assumed to be normally distributed. Without this or some similar assumption further progress cannot be made. To persons to whom a set of test scores is merely a series of numbers obtained to provide material for algebraic manipulation, this book will be a convenient treasure-house of equations and formulas—hundreds of them. Psychological research workers will, however, one feels, turn aside with feelings akin to horror. E.G.C.

The Changing Culture of a Factory. By Elliott Jaques, with an Introduction by A. T. M. Wilson. London: Tavistock Publications Ltd. 1951. Pp. xvii + 341. 28*s*.

This is a report of a three-year study carried out by the Tavistock Institute of Human Relations in a London engineering factory. It begins with a section on background and historical material (Part I) which leads to a description of five independent case studies (Part II) and is rounded off with an analysis and interpretation of the main trends (Part III). The project was based on field theory which 'holds that a particular event may best be understood as the outcome of other interacting events in the larger field in which it occurs' rather than in terms of simple cause and effect. In particular, it represents an attempt to adapt some of the principles and methods of psycho-analysis to the study of group behaviour and in so doing it makes use of concepts such as transference, resistance and unconscious forces.

The Tavistock Research Team, with Dr Jaques as Director, was engaged for the most part in studying the problems and difficulties of various factory committees. The usual procedure was for a member of the Team to attend all committee meetings as observer and consultant. From time to time he drew the attention of the members to unrecognized difficulties and resistances. In this way, the committee was able 'to unearth and identify some of the less obvious influences affecting its behaviour' and thus 'to take better decisions concerning changes they needed to make'. It is also claimed that 'once a group has developed insight and skill in recognizing forces related to status, prestige, security, authority, suspicion, hostility and memories of past events, these forces no longer colour subsequent discussion nor impede progress to the same extent as before'.

After reading this book, one is left with the uneasy feeling that any attempt to assess its merits and defects is likely to be influenced by unconscious and irrational factors. Yet despite this haunting fear, it seems fairly certain that the factory was fortunate in having the help of Dr Jaques and his colleagues; a belief that receives some support by their continued employment long after this book was written. Although their activities were confined to a single factory, the book contains much that will interest the democratically minded employer who is willing to share responsibility and is anxious to establish good relations in and between different factory groups. Whether it will have much influence on those employers who are content to jog along as best they can is rather doubtful, especially when they are told that 'in the process of studying a particular problem, new problems are created' and 'the resolution of stresses within a group is to be regarded as a continuous and endless task'. It is also important to remember that the detention, unravelling and treatment of unconscious forces in group behaviour is a job for the skilled social practioner; it needs knowledge, insight and tact and an unskilled person may easily do more harm than good.

Although in this type of research each problem must be treated on its merits, it is possible that a series of independent case studies will reveal recurring patterns and lead to generalizations of some importance. The validity of these generalizations will, however, depend very largely on the methods employed in the

llection and treatment of the data on which they are based and it is this aspect of the investigation hich gives rise to doubt. Thus, in each of the five case studies, the research consultant had to deal with highly complex situation made up of a number of known and unknown variables. It seems highly obable that his method of 'working through' the situation with the people concerned would have fferent effects on different factors and the final outcome would be the resultant of these effects. Further, is quite impossible to determine how far the rate of progress and the subsequent decisions were due to e consultant or to the members of the committee. They got results but no one can say just how they got em. It is, of course, quite true that merely 'to identify and define the main forces concerned with social haviour is a first and necessary step in quantification', but unless a field theory approach will allow r Jaques and his associates to isolate and measure the effects of each of the identifiable factors while eping the remainder reasonably constant their findings are likely to be of limited value and application. dmittedly the present book is described as a progress report and is 'not intended to be a statement of finite conclusions'. It describes the events leading up to a change but not the effects of the change, and ere is no means of knowing whether the effects were good or bad. Although follow-up studies are said be in progress and other reports are promised, it might have been better to defer publication until the sults were available for assessment.

The book as it stands has many good features and the student of human relations will find it both teresting and suggestive. A field theory approach undoubtedly has definite attractions and opens up w and important possibilities. If the field theorists are willing to devote a little more of their time and ility to 'working through' their methods they should be able to fill a gap in the field of social research.

S.W.

sychology of Human Relations in Industry. By FRANZISKA BAUMGARTEN. London: Pitmans 1951. Pp. x+205. 15*s.* net.

his book, by the general secretary of the International Association of Applied Psychology, is a transla-on of the second edition of *Die Psychologie der Menschenbehandlung* published in 1946 in Zurich. After 11-page historical survey of the growth of interest in industrial relations and a chapter on 'indirect' ethods of improving these relations (improvement of living and working conditions and the like), the thor settles down to her main theme which is 'How to Improve Personal Relations in Industry—rticularly those between the boss and the worker?' The answer to this question is told in a series of rmons, most of which are addressed to the boss. He is told that employees differ in their make-up and ould therefore be treated differently—but impartially. Examples are given. He is bidden to be—or to mulate being—a good example to subordinates in general behaviour; he is warned, again and again, gainst losing his temper. More examples are given. Next, the 'bad' worker is urged to be less suspicious, ss defiant and more co-operative. No one could possibly disagree with this main section of the book. It full of wise saws which have not been in dispute these last two or three thousand years. Characteristi-lly enough, the text is peppered with quotations from Chinese philosophers, and the flavour of the book ould put it alongside Pascal's *Pensées* and Marcus Aurelius' *Meditations* rather than on the shelf of merican and British works on industrial psychology. But modern instances, anyhow for a British ition, are harder to find. The N.I.I.P. is mentioned only once—an article in its Journal by F. C. rtlett, 1926. Neither Mayo nor Roethlisberger appears in text or index. And for a book intended 'for nployers, executives, and welfare officers', the bibliography needs drastic revision. Out of 95 books and ticles listed, 26 are in German and 24 in French. Only one of the whole 95 is reasonably accessible e. published in Britain) *and* recent (i.e. in the last 20 years).

D. MCMAHON

sychoanalysis and Politics. By R. E. MONEY-KYRLE. London: Duckworth, 1951. Pp. 183. 9*s.* net.

his is a disarming little book. From its title one might expect something rather heavy with a shadowy *keit* or *-ismus* at the tail of every other word; but it is concrete, shrewd and readable. If one were to iticize it in terms of science or philosophy it would not stand up very well. Sweeping generalizations op up all over the place. For example 'Moral behaviour may be defined as behaviour dictated by the ar of a sense of guilt' (p. 54). Or, as a footnote: 'Religion is a form of psychotherapy which promotes belief in the existence of idealized good objects as a defence against persecutory and depressive guilt' . 84).

The main thesis is based upon the division of a number of Germans examined after the last War into thoritarians with persecutory consciences, and humanitarians with depressive consciences. These two pes are related first to the early family situation, and then to the State. It is shown in terms of psycho-

analytic theory how the differences arise, and it is also shown in psychological rather than ethical terms that the humanitarian is a better human being.

This bald outline does less than justice to the interest and value of the book. Brevity requires dogmatism. It is perhaps unfortunate, however, that a book intended for the ordinary reader should contain so many irrelevant bits and pieces of Freudian theory presented in a way which makes rejection easy. The trouble is, I suppose, that everyone of us writing a book for the general public sees a professional colleague peeping over his shoulder. J. DREVER

The Health of the Mind. By J. R. REES. London: Faber and Faber. 1951. Pp. 208. 9*s.* 6*d.*

Dr Rees, in his foreword to this new edition of his book, says that, on re-reading it after twenty years, *mirabile dictu*, he actually agreed with most of it. There is, in fact, nothing surprising about this, as the book is a fine résumé of the generally accepted fundamentals of mental health, which have stood the test of years and experience. It is meant for the ordinary reader, and is therefore simple and clear in style: this simplicity and clarity do not detract in any way from its soundness and truth.

It is a sympathetically written account, first of physical and psychological mechanisms involved in our social and emotional life, and then of the way in which these mechanisms can go wrong and break down. There is also some sound advice of a kind which will appeal to the layman by its practical applicability and to the psychologist by its width of view and genuine understanding. A. M. SILVER

The Psychology of Adolescence. By JOHN E. HORROCKS. Boston, U.S.A.: Houghton, Mifflin and Co. 1951. Pp. xxvi+614. $4.50 net.

This is an American book concerned with American Youth. Until recently a book on adolescence would be considered to have general application whatever the country of its origin; now all is changed and Dr Horrocks would be the first to accept such a limitation to his field. In the formation of attitude and interest, the American Way of Life can be as significant a factor as the onset of puberty. 'Superman', family autos, dates, baseball, the radio (television has not yet yielded its spate of research) are, in the main, peculiar to one culture and untranslatable into other patterns. Nor is the story yet finished; the publication of the Kinsey research itself will doubtless have its effect in the future, if it is not already doing so, on the teenager, who is so very concerned to 'achieve the norm'. Most of the terms peculiar to this culture are explained, much of the writing is vivid—'the typical adolescent boy. . .is a fugitive from the term "cissy",'—and when it is statistical and quantitative, it is clear and fully documented. The author has not been quite able to separate fact and comment, and although he does not claim to deal with morals, the expressions 'unfortunately', 'it is better', etc. creep in, as indeed they are almost bound to when the upbringing of the young is concerned.

Here is no attempt to regard the adolescent as *sui generis;* he is shown as being of the same stuff as the child and the adult, and no youth worker, on reading this book, will get the misleading impression that adolescence is inevitably a time of strain and stress and the teenager, an unpredictable person, best herded into youth clubs so that he can be dealt with in isolation from the rest of his community. His physical and his psychological development are well documented in all their variety. He is considered as an individual, but much more as a member of his family and of his social group. He is shown to have extremely varied interests, attitudes that can be measured and media of communication that are varied and bewildering.

With the caveat that we made in our opening paragraph, this is a book that can be widely recommended to all workers in the field of adolescence. J. M. GIBBS

The Sociology of the Patient. By EARL LOMON KOOS. N.Y., Toronto and London: McGraw Hill. 1950. Pp. xviii+266. $3.00 net.

This book is a fascinating and readable introduction to sociological principles and their application in human relationships and in various fields of everyday activity and behaviour. The author traces for the young student a simple outline picture of the evolution of man as a Primate with special emphasis on those aspects of man's equipment which make the human being unique. The opening chapters thus stress the importance of the individuality of man which is a helpful corrective to some modern tendencies to think of human beings only as members of groups, or even crowds.

The development and significance of the many and varied groups and culture patterns in modern life, including some helpful descriptions of racial and national attitudes, are simply and clearly outlined. The

apters on the changing position and responsibilities of the family will be found most helpful to readers ncerned with social problems, notably those arising from the 'broken home', whether this be due economic or matrimonial difficulties.

The statistics provided in relation to the development of industry, the decline of rural communities th some of their special characteristics, and the growth of urbanization, are not in themselves relevant similar trends in Great Britain. But the principles which these figures illustrate are clearly set out, and n be applied in any community. In this connexion the questions at the end of chapters 7, 10, 12, and are particularly useful; and the answers to these, which are within the capacity of any nurse or social rker to obtain for her own area, should be most enlightening.

The sociological facts which should be at the disposition of any intelligent and conscientious nurse, ial worker, or teacher, and the methods to be employed in securing these, are clearly presented in this lume. In addition, however, perhaps its most valuable contribution to students, nurses, and other rkers in the field of human relationships, is the implied insistence throughout on seeing their 'cases' as ing individual men and women dynamically affected by the social pattern of the various groups, be se economic, familial, recreational, or religious, which are their background. NORAH MACKENZIE

nführung in die moderne Psychologie. By WALTER TOMAN. Vienna and Stuttgart: Humboldt. 1951. Pp. 268.

this country this little book might be treated as a gift for would-be readers of psychological German o need practice and whose spirits have not yet been broken by separable prefixes. The thinking and the rman in which it finds expression are lucid and to the point, the book is small enough to slip into ucksack and to offer to the linguistically humble the possibility of having *finished* it within a relatively ort time. It has a German-English glossary of technical terms.

The book is published in the German and Austrian series which is roughly equivalent to our Home iversity Library. It would be fair to say that it is one thing to understand the German and another to derstand and digest what the author has to say. For he has packed the history of psychology (Plato Lewin and others) into chapter 1, the main principles of theoretical systems such as behaviourism, ycho-analysis, topological psychology and multiple factor theory into chapter 2. He then has gone on consider general psychology, clinical psychology and social psychology, he adds a discussion of the pe of psychology and finishes with a good stiff dose of statistical methods.

All this has much to commend it, but two warnings should be given. First, though the author's references British and American literature are mainly excellent, there is the slight suggestion that he is continental tending to leave to 'psychotechnics' areas which in England would fall within respectable occupa- nal psychology. Secondly, to the English reader, schooled to believe that it is more fun to think than know, the book savours a little too much of the cram volume which hard pressed people might be apted to learn by heart for examination purposes. This is a peril, for lucid and succinct as is much of Toman's exposition, it seems at least to the present reviewer that parts of the book would be largely ntelligible unless one had browsed on the material already in a much more leisurely fashion.

mo Faber. A Study of Man's Mental Evolution. By G. N. M. TYRRELL. London: Methuen and Co. Ltd. 1951. Pp. xii + 206. 15*s.*

Tyrrell is well known for his work in psychical research. In this book he turns his attention to the blem of mind and its relation to the objective world. He was led to do so because his work in psychical earch suggests that in many cases where *A* is supposed to have acquired, by non-sensory means, ormation which *B* possessed, he had really got information from a source 'outside the spatial world and yond the reach of experimental control' (p. 38). This led him to consider whether those scientists who tinately refuse to consider this possibility are not perhaps limited in outlook. The scientist, he argues, nes to his study of the universe with his assumptions ready made. He accepts only certain types of planation: and consequently there is no reason to think that he can explain all there is. Man's mind has n adapted to serve practical purposes and not to see the puzzling aspects of truth. The practical ellect (*Homo faber*) lives within a well-rounded sphere set by its own limitations. Psychical research ruptured the envelope. Man can rise to wisdom, become *homo sapiens*, if he will make the effort. In realm that is not subject to the practical intellect, verbal clarity cannot be obtained and metaphor st be used to express insights.

This thesis is worked out with reference to difficulties in biology and psychology. The conclusion, which ot altogether what one would expect, is that philosophy exists not to solve problems but to throw them stereoscopic relief. W.H.F.B.

OTHER PUBLICATIONS RECEIVED

Crow, Lester D. and Crow, Alice. *Mental Hygiene.* New York and London: McG Hill. Second Edition. 1951. Pp. xviii + 433. 38*s.* 6*d.* net.

Fulton, J. T. *Experiment in Dental Care.* Geneva: World Health Organization, M graph Series, No. 4. 1951. Pp. 87. 5*s.* net.

Church of England Moral Welfare Council. *The Chaplain.* London: Ch Information Board. 1951. Pp. 13. 1*s.* net.

Committee appointed by the Vice-Chancellor. *Research and Scholarship in University of Sydney.* Sydney. Australasian Medical Publishing Company Lim 1951. Pp. 120.